University of Kent

PhD Thesis:

Experimentally validated computational docking to characterize protein- protein interactions

Supervisor:

Prof Vadim Sumbayev

Dr Luca Varani

Candidate:

Elsa Livoti

Acknowledgements

I would like to thank my research advisor Luca Varani and my supervisor at the University of Kent, Prof Vadim Sumbayev, for giving me the opportunity to do the PhD about such interesting projects and for guidance, constant availability, support and advice.

My gratitude goes also to the people sharing the lab with me: Luca, Mattia, Marco and Daniela.

A special thanks to Luca and Mattia for supporting me in the lab activities and in the difficult moments of my years spent at the IRB Institute.

List of Pubblications

Epitope mapping by solution NMR spectroscopy.

M.Bardelli*, E. Livoti*, L. Simonelli, M. Pedotti, A. Moraes, A.P. Valente and L. Varani

* Equal contribution

Journal of Molecular Recognition 28 July 2014

Rational engineering of a human anti-Dengue antibody through experimentally validated computational docking.

Luca Simonelli, Mattia Pedotti, Martina Beltramello, Elsa Livoti, Luigi Calzolai, Federica Sallusto, Antonio Lanzavecchia and Luca Varani

PLoS One, February 6, 2013

HMGB1 promotes recruitment of inflammatory cells to damaged tissues by forming a complex with CXCL12 and signaling via CXCR4.

Schiraldi M, Raucci A, Muñoz LM, Livoti E, Celona B, Venereau E, Apuzzo T, De Marchis F, Pedotti M, Bachi A, Thelen M, Varani L, Mellado M, Proudfoot A, Bianchi ME, Uguccioni M.

J Exp Med. 2012 Mar 12;209(3):551-63. Epub 2012 Feb 27

Computational docking of antibody-antigen complexes, opportunities and pitfalls illustrated by influenza hemagglutinin.

Pedotti M, Simonelli L, Livoti E, Varani L.

Int J Mol Sci. 2011 Jan 5;12(1):226-51.

A blue light-inducible phosphodiesterase activity in the cyanobacterium Synechococcus elongatus.

Cao Z, Livoti E, Losi A, Gärtner W.

Photochem Photobiol. 2010 May-Jun;86(3):606-11. Epub 2010 Apr 7.

Interdomain signalling in the blue-light sensing and GTP-binding protein YtvA: a mutagenesis study uncovering the importance of specific protein sites.

Tang Y, Cao Z, Livoti E, Krauss U, Jaeger KE, Gärtner W, Losi A. Photochem Photobiol Sci. 2010 Jan;9(1):47-56. Epub 2009 Nov 30.

Figures

The images concerning the third section of the thesis have been prepared by myself. The software Pymol was used for three-dimensional structure visualization; Sparky for NMR figures; SPR images were generated with the BioRad Proteon software; scoring plots were generated as output of the docking simulations using Rosetta Dock; all the other figures have been created by myself with graphic software. Dr Luca Simonelli and Dr Mattia Pedotti helped with figure generation during the co-work on the research projects.

Abbreviations:

HSQC: Heteronuclear Single Quantum Coherence
HMGB1: High Motility Group Box 1
NMR: Nuclear Magnetic Resonance
SPR: Surface Plasmon Resonance

INDEX

Abstract

Each biomolecule in a living organism needs to adopt a specific three-dimensional conformation to function properly. Function itself is usually achieved by specific interactions between biomolecular units. Structural knowledge at atomic level of biomolecules and their interaction is important to understand the mechanisms leading to biological response and to develop strategies to interfere with them when necessary.

Antibodies are molecules of the immune system playing an ever more prominent role in basic research as well as in the biotechnology and pharmaceutical sectors. Characterizing their region of interaction with other proteins (epitopes) is useful for purposes ranging from molecular biology research to vaccine design.

During my PhD studies I used a combination of solution NMR mapping, molecular biology and computational docking to provide a structural and biophysical characterization of new neutralizing antibodies from Dengue virus recovered subjects, comparing the binding of the same antibody to the four Dengue serotypes and the binding of different antibodies to the same serotype. We were able to rationally mutate an antibody to first alter its selectivity for different viral strains and then increase its neutralization by ~40 folds. For the first time, this was achieved without the availability of an x-ray structure.

In a second sub-project, I investigated the interaction of the chemokine CXCL12 with the chromatin-associated protein HMGB1, confirming their direct interaction (only proposed but never proved before) and providing a

structural explanation for the HMGB1 dependent increase of CXCL12 cellular activity. High profile publications resulted from the two above projects.

The above mentioned projects relied heavily on solution NMR spectroscopy, which is ideally suited to the atomic level characterization of intermolecular interfaces and, as a consequence, to antibody epitope discovery. Having provided a residue-level description of a protein-protein interface by NMR, we subsequently used this experimental information to guide and validate computational docking experiments aimed at providing a three dimensional structure of the protein-protein (or antibody-protein) complex of interest. In collaboration with other members of my research group I validated the use of NMR and computational simulations to study antibody-antigen interactions, publishing two reviews in collaboration with other members of my research group.

Section I

Epitope mapping by

NMR

Abstract

Antibodies play an ever more prominent role in basic research as well as in the biotechnology and pharmaceutical sectors. Characterizing their epitopes, i.e. the region that they recognize on their target molecule, is useful for purposes ranging from molecular biology research to vaccine design and intellectual property protection.

Solution NMR spectroscopy is ideally suited to the atomic level characterization of intermolecular interfaces and, as a consequence, to epitope discovery. Here it's illustrated how NMR epitope mapping can be used to rapidly and accurately determine protein antigen epitopes. The basic concept is that differences in the NMR signal of an antigen free or bound by an antibody will identify epitope residues. NMR epitope mapping provides more detailed information than mutagenesis or peptide mapping and can be much more rapid than x-ray crystallography. Advantages and drawbacks of this technique are discussed together with practical consideration.

Introduction

Antibodies are proteins capable of recognizing their target molecule, called antigen, with high affinity and specificity. They play an ever more prominent role in basic research as well as in the biotechnology and pharmaceutical sectors[1]. Antigens are recognized through atomic interactions between up to six highly variable antibody loops, called antigen binding loops, and the so-called epitope, which is the region of the antigen contacted by the antibody.

Generally speaking, characterizing epitopes will improve our understanding of the molecular basis for antibody-antigen recognition. This in turn could allow, in the long run, altering existing antibodies or designing new ones with desired properties. Discovering the epitope of a given antibody also has more short term purposes. If antibody binding affects a protein function, then the epitope may indicate which part of the protein is responsible for such function. The epitope of an antibody capable of blocking the binding of a protein to cells, for instance, may reveal the receptor binding site of the protein antigen. Furthermore, including a detailed epitope description in a patent is expected to strengthen the protection of intellectual property[2]. Epitope discovery is also highly relevant for vaccine design. Most vaccines act by injecting an antigen in an individual and generating an antibody response. Using full pathogens, e.g. a virus, as vaccines has two main disadvantages: the pathogen needs to be suitably inactivated; the full pathogen may have regions that generate ineffective antibodies. For instance, the seasonal influenza vaccine contains a highly variable region that generates antibodies incapable of protecting against future circulating influenza strains. One strategy to overcome these problems is to look for highly effective antibodies[3], find out what their epitope is and finally use the epitope, rather than the full pathogen, as vaccination agent.

Mapping an antibody epitope

Since the epitope is the region of an antigen that interacts with the antibody, a three-dimensional structure of an antibody/antigen complex, usually obtained by x-ray crystallography, provides the most thorough description of epitopes. An atomic structure, in fact, allows the observation of specific intermolecular contacts.

X-ray structures, however, are often laborious to obtain and not always available. On the other hand, epitope prediction by computational algorithms is fast and affordable but has not yet reached widespread accuracy. In the case of protein antigens, other techniques like peptide mapping and site directed mutagenesis provide important clues to an antibody epitope but have significant limitations. In peptide epitope mapping short portions of a protein sequence are synthesized and then tested for binding to an antibody. A positive result indicates that the peptide sequence is indeed part of the antibody epitope. Antibody recognition, however, often requires the presence of a well-defined three-dimensional structure in the epitope (so-called conformational epitopes). It is often not trivial to reproduce such structure in a peptide library. In these circumstances it is impossible to know if a negative result implies that the peptide sequence is not part of the epitope or that, instead, the proper atomic structure is missing. Finally, if mutagenesis of single protein residues prevents antibody binding, then it may be assumed that such residue is part of the epitope, although allosteric effects need to be taken into consideration. The main problem with this approach is that it is difficult to obtain information on all residues forming the epitope.

Here it’s described how solution NMR spectroscopy can be used to characterize antibody epitopes. The strategy is faster than x-ray crystallography, is effective for conformational epitopes and provides a more thorough description than peptide mapping or site directed mutagenesis. Advantages and disadvantages are illustrated below, together with some practical considerations. Although the description focuses on protein antigens, similar considerations are valid for other molecules.

NMR epitope mapping

Solution NMR spectroscopy is particularly suited to the characterization of intermolecular interactions, including the determination of antibody epitopes. The NMR signal, in fact, is exquisitely sensitive to the local chemical environment. When an intermolecular complex forms, the chemical environment of interface atoms changes as they get close to the atoms of the interacting partner, and so does their NMR signal. By comparing the NMR spectrum before and after complex formations, interface atoms can be individuated due to changes in their NMR signal (figure 1).

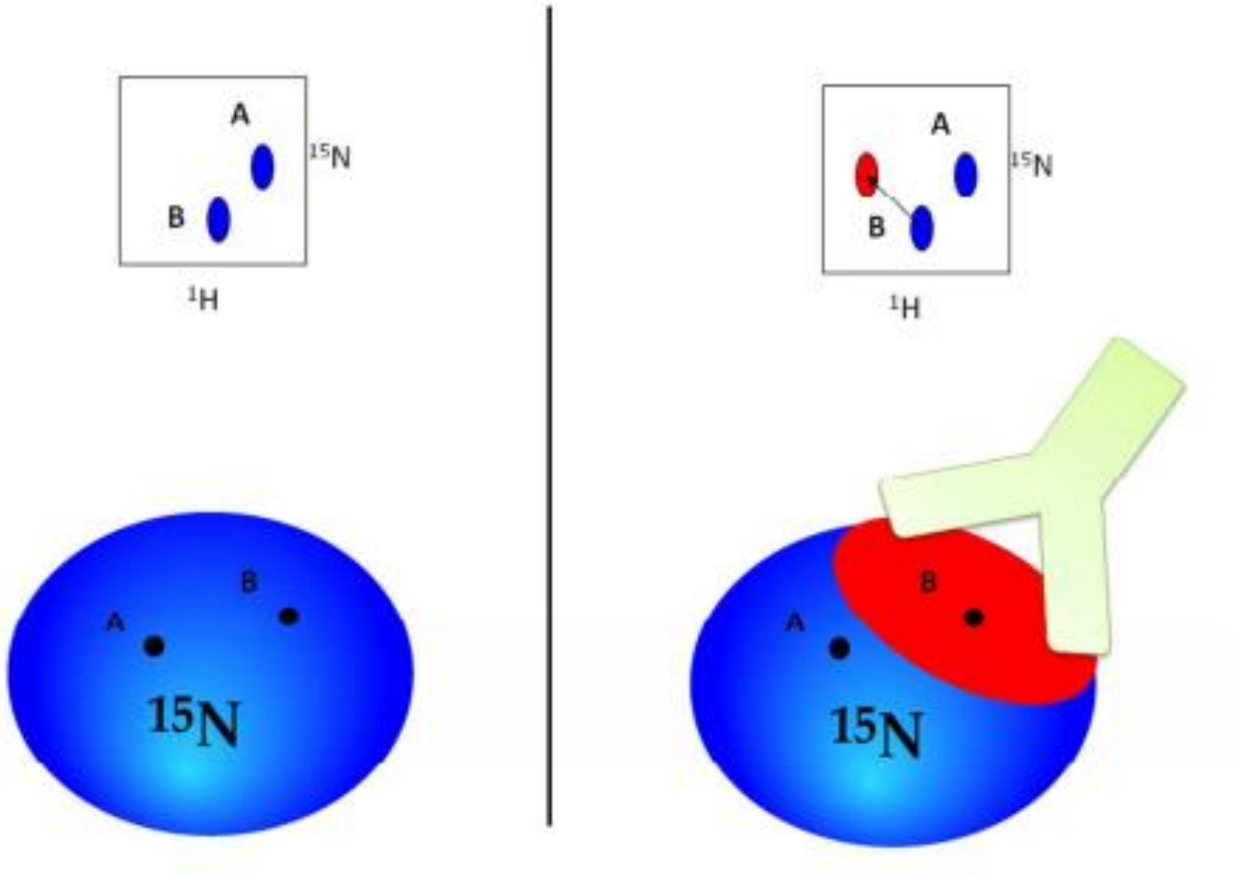

Figure 1: Schematic representation of an NMR epitope mapping experiment. Residues A and B of a ^{15}N labelled protein antigen (left) generate a single peak in a ^{15}N-HSQC experiment, schematically shown at the top. Upon antibody binding (right), epitope residues experience a different chemical environment and their NMR signal changes, as shown for residue B in the scheme. The antibody is unlabeled and its signals do not appear in ^{15}N-HSQCs. By comparing the NMR spectrum of the antigen free and in complex with the antibody, epitope residues whose signal is affected by binding can be identified.

Here we illustrate the characterization of protein epitopes for simplicity, but similar considerations are valid for other molecules.

In a ^{15}N-HSQC experiment[4] the backbone NH group of each protein residue (except prolines) generates a single NMR signal (figure 2).

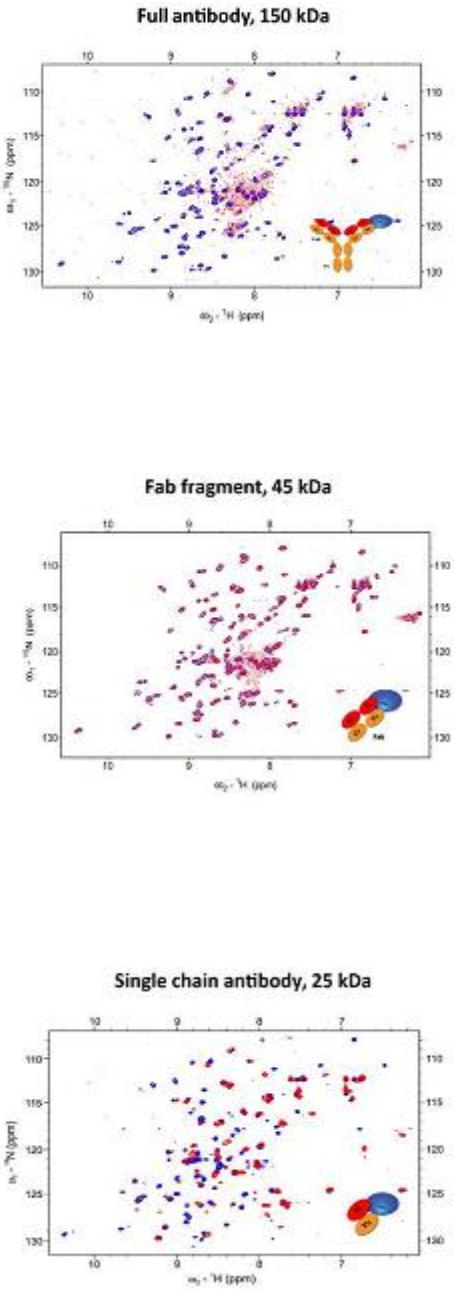

Figure 2: ^{15}N-HSQC spectra of an antigen derived from Dengue virus either free (blue) or in complex (red) with a full antibody (left), with a Fab fragment

(center) or with a scFv fragment (right). A schematic representation of the molecules involved is shown in the lower right corner. Each peak correspond to a single antigen residue in these experiments. Signal intensity increases and broadening decreases when analyzing smaller antibody fragments. Changes in the position of some peaks upon complex formation can be discerned.

The position of these signals is so sensitive to the protein conformation that ^{15}N-HSQC spectra are often referred to as protein fingerprints. Protein labelling with ^{15}N (see below) is required because the naturally occurring nitrogen isotope, ^{14}N, gives complicated NMR spectra (I=1).

The idea behind NMR epitope mapping, schematically illustrated in figure 1, is to:

1) Record a spectrum of the free, labeled protein antigen.

2) Add unlabeled antibody and record a HSQC of the complex. Since the antibody is not labelled, its residues do not produce an NMR signal and thus do not appear in the spectrum.

3) Overlay and compare the spectra of the antigen free and in complex with the antibody. If a signal changes position it means that the chemical environment of the residue generating it has changed; in other words, the residue is affected by antibody binding and is likely to belong to the epitope.

The list of all antigen residues whose signal changes upon antibody addition provides a full description of the epitope. ^{13}C-HSQCs may be used instead of ^{15}N-HSQC. However, since the former has one peak for each CH group in the labelled molecule and since there are many more CHs than NHs in a protein,

^{13}C-HSQCs are often marred by spectral overlap and are consequently less informative. In case of spectral overlap, in fact, it is often difficult to discern if peaks change position upon complex formation and to distinguish which of the residues with overlapping signal is affected. Due to similar considerations, ^{15}N-HSQC spectra of proteins with many residues are more difficult to analyze. Generally speaking, because of spectral overlap proteins larger than 200 residues may not be suitable to NMR epitope mapping without selective labeling. This problem can be alleviated by labelling only specific amino acids in the protein, for instance all tyrosines. This is usually achieved by including appropriately labelled nutrients in the growing media during protein expression[5,6]. Alternatively, protein segments of a few contiguous residues can be selectively labelled[7-9]. Large proteins present a second challenge to NMR epitope mapping. The intensity of the NMR signal, in fact, decreases at increasing molecular weight[5,10-12]. The signal arising from large proteins (or protein complexes) may be undetectable. However, TROSY versions of HSQC experiments can alleviate the problem and provide a signal even for large molecules[13]. The decision to run simple HSQCs or their TROSY equivalent should be taken on a sample by sample basis, since one or the other may offer better signal intensity according to specific sample conditions. Protein labelling with deuterium (2H), which may complicate sample preparation, is required for optimum TROSY performance. When conducting NMR epitope mapping studies we typically acquire both a quick HSQC and TROSY experiment, evaluate which is better and then run a long experiment with the chosen option. Both spectral overlap and the lack of sensitivity at increasing molecular weight can be alleviated by the use of high field NMR

spectrometers commonly available today. We have successfully performed NMR epitope mapping experiments on a 600MHz spectrometer in some cases, but 800MHz or more may be necessary and are certainly beneficiary.

Which is which? Protein assignment

Comparison of ^{15}N-HSQC experiments of a protein antigen free and in complex with an antibody identifies NMR signals affected by antibody binding and therefore likely to belong to residues that are part of the epitope. In order to map the epitope on the protein primary sequence or three-dimensional structure we need to know which signal belongs to which residue. A detailed description of this process, called protein assignment, goes beyond the scope of this thesis[14-16]. Suffice to say that assignment requires collection and analysis of several different NMR experiments, usually obtained from proteins labelled with both ^{15}N and ^{13}C. Assignment may take as little as a week, as long as a few years or be altogether impossible to obtain. Size of the protein, sample behaviour (e.g. propensity to aggregate, lack of stability over time etc) and protein secondary structure may all affect the difficulty and thus the time required to assign a protein. Although each case is different, proteins with less than 120 residues are often trivial to assign and proteins above 200 residues become complicated and require longer time and/or strategies like selective labelling. An unsupervised search of the Protein Data Bank shows a total of 10996 structure determined by NMR spectroscopy (figure 3).

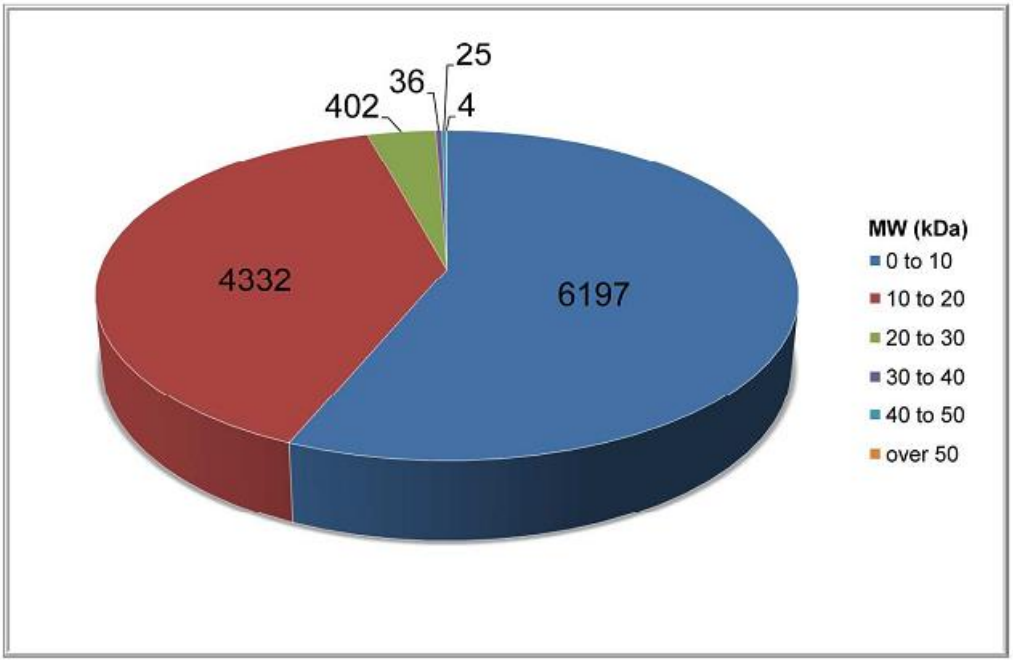

Figure 3: Distribution of molecular weights of the NMR structures deposited in the PDB. Colors indicate molecular weight ranges (kDa) as shown in the legend. The number of structures for any given rage is indicated in the chart.

The vast majority of these have molecular weight below 20kDa and only 29 of them above 40kDa. It is important to note that structure determination is much more complicated than assignment alone[17,18]; nonetheless, the data provides an idea of what has been achieved so far with NMR.

Assignment of the free antigen is essential but formal NMR assignment of the antigen in complex with the antibody, which is much more difficult to obtain, is not strictly necessary unless a quantitative analysis is desired (see below). If the signal of a given residue changes position (or disappears) upon complex formation, then we know that the residue is affected by antibody binding even if we do not know its exact position in the spectrum of the complex. In other words, it is not necessary to know where a peak moves upon complex formation, it suffices to know that it is not in the position it had in the

spectrum of the free antigen. Complex assignments can be estimated by simple comparison to the free antigen spectra. When changes are small it is usually trivial to pair each peak in the complex to its free counterpart. The larger and more numerous the changes, the less reliable this approach becomes. If a protein changes its structure upon complex formation, for instance, it may be impossible to assign it by comparison. If complex assignment is only available by comparison to the free spectra and ambiguities are present, one possibility is to consider the "minimum shift difference"[19]. The amount of chemical shift change is defined as the difference between a peak in the free spectrum and its closest peak in the bound state. This guarantees that the entity of the chemical shift change is equal or larger to the one set in the analysis, but it is not overestimated. Once assignment is available, directly or from a public database[20], epitope mapping of any antibody interacting with the same antigen can be obtained in a matter of hours. HSQC experiments on the antigen free and in complex with one antibody require 24 hours at most. Comparison of the two spectra and identification of the NMR signals affected by antibody binding can be completed in a couple of days. Having the assignment, mapping those signals to the protein sequence and to the protein structure is instantaneous.

What is a significant change in NMR signal? Define residues affected by antibody binding

We can identify epitope residues by detecting changes in their NMR signal upon formation of the complex with an antibody. Although the concept is simple, a precise definition of what is considered a significant change might not be. How much does an NMR peak need to change position to be considered different is a question open to subjective interpretations. When analyzing quantitatively a ^{15}N-HSQCs, one possibility is to measure the distance between a peak position in the free and bound form in both the nitrogen and proton dimension and combine the two terms in a formula like $[(\Delta NH)^2+(\Delta^{15}N/10)^2]^{1/2}$ (different formulas may be used). This generates a "chemical shift perturbation index"-CSP (the term "chemical shift" indicates the position of the NMR signal). The difference in ^{15}N chemical shift is divided by 10 to compensate the larger frequency range. It is then possible to average the values off all antigen residues and consider the signal of a given residue significantly different upon complex formation if its chemical shift perturbation index is more than one standard deviation away from the average. This method may work well for small, well behaved protein complexes. In our experience, however, it is often not suitable to large antibody/antigen complexes marred by spectral overlap, signal broadening and other complications. One alternative is to plot the chemical shift perturbation index of each antigen residue in a chart like that shown in figure 4.

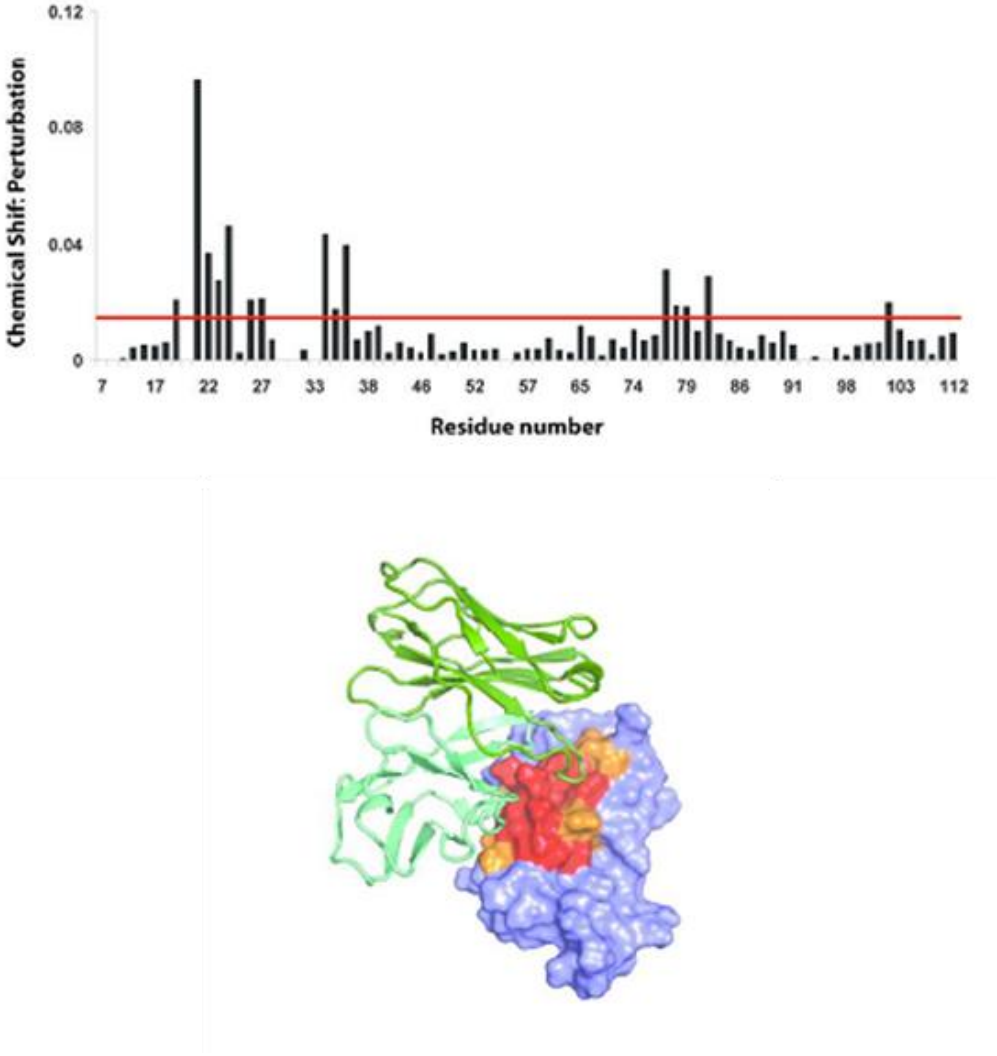

Figure 4: Quantitative analysis of an NMR epitope mapping experiment. The Chemical Shift Perturbation index (CSP, see main text) is plotted for each antigen residue. Regions with higher CSP are identified and an arbitrary threshold is set (red line). Antigen residues with CSP above the threshold are considered to be affected by antibody binding. On the bottom, these residues are colored in red and orange on the structure of an antigen (blue) in complex with the antibody (green). They define an epitope region compatible with the dimension of an antibody binding site.

Similar plots often reveal the presence of contiguous residues whose signal is affected by complex formation and often allow setting an arbitrary threshold: if a residue has a perturbation value above the threshold it is considered affected by antibody binding. See below for further details on how to set the threshold. Due to the complexity of the system, we prefer to adopt a more qualitative approach whose flexibility offers better results in antibody/antigen complexes, in our experience. We divide the ^{15}N-HSQC peaks, and thus the protein residues, in four classes: 1) large changes: residues whose position shifts by more than 0.4ppm in the nitrogen or 0.04ppm in the proton dimension and that are clearly visible in the NMR spectrum (no overlap or other uncertainties); slightly different threshold values might be set on a case by case basis.

2) Small changes: residues whose signal changes by less than the values above.

3) No changes: residues whose NMR signal does not change upon complex formation.

4) Unknown: the effect of antibody binding on these residues cannot be determined due to lack of assignment, spectral overlap or other problems. Figure 5 shows example for each class.

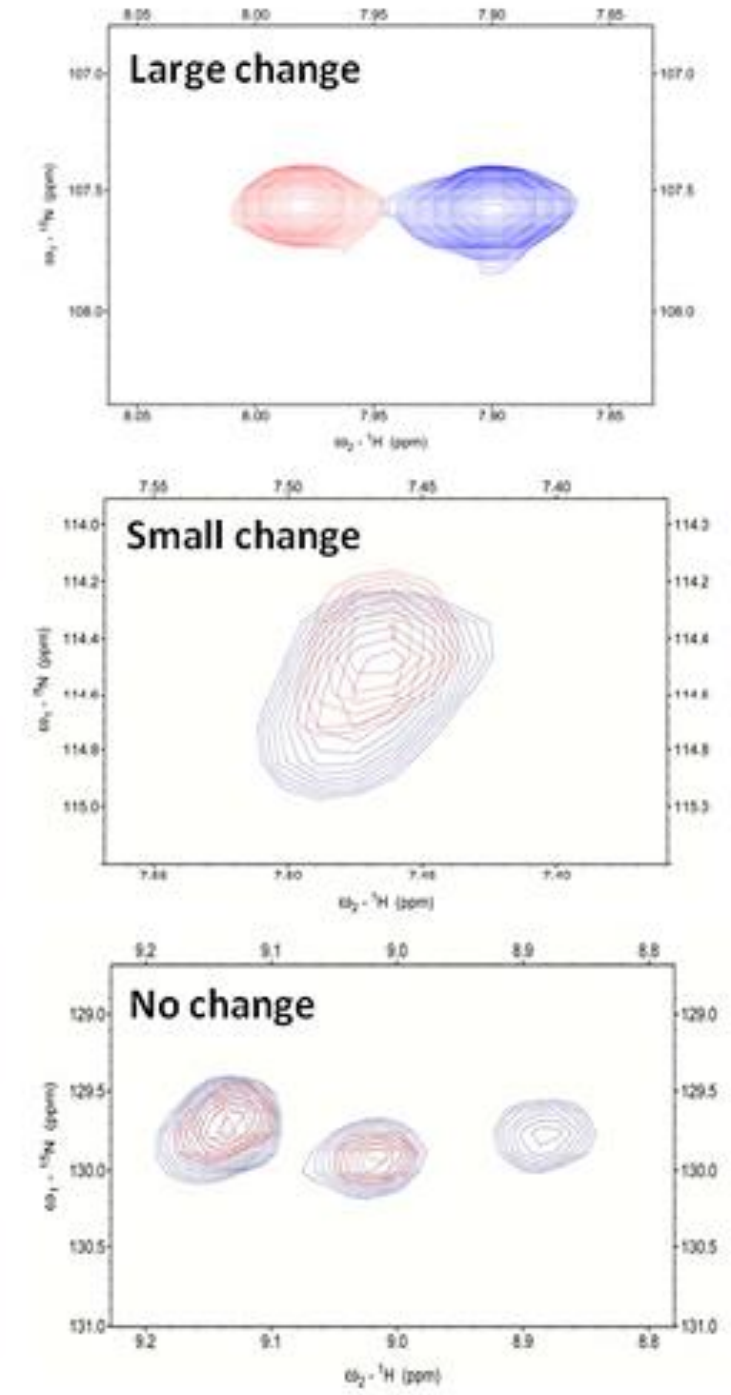

Figure 5: 15N-HSQC peaks of an antigen free (blue) or in complex with an antibody (red). Differences between free and bound spectra indicate that the residues generating the peaks shown are affected by antibody binding and likely part of the epitope. Examples of large, small or no changes upon complex formation are shown.

It should be noted that the magnitude of the NMR change (large/small) is not directly correlated to the distance of a given residue to the antibody. Although the effect is not detected at distances larger than approximately 7Å, a change in NMR signal depends on the magnetic properties of the local environment surrounding a given chemical group. There is no simple equation correlating the amount of change in the NMR signal to the distance between two residues on interacting molecules. Proximity of an aromatic ring, for example, generally provokes larger changes than equal distance to a non-aromatic ring due to ring current effects having a strong impact on the nearby magnetic properties[21]. Having divided the antigen residues in the above four classes, we map those with "large changes" on the antigen structure. They usually form a well-defined, contiguous region compatible with the known dimensions of antibody binding sites. Residues in the "small changes" class and residues that apparently have large changes but are marred by some uncertainty like partial spectral overlap are included only at a second stage and only if there is clear structural evidence that they are part of the NMR derived epitope. For instance, a "small change" residue surrounded (in the three-dimensional structure) by "large changes" is usually added to the epitope. We treat "small change" residues on the fringe of the epitope defined by the large changes with caution and include them only if there is other evidence justifying it, for instance from mutagenesis or computational simulations. Isolated "small changes" residues may appear on regions of the protein clearly outside the well-defined epitope, for example on the face of the protein antigen opposite the epitope. In such cases they are not further included in the structural analysis. Finally, "unknown" residues are never included in the NMR derived

epitope, not even if they are in the middle of a clearly defined epitope. Although the initial analysis should strictly rely on undisputed NMR information to avoid bias, the process becomes iterative and structural considerations are used to help resolve ambiguities in the NMR spectra. A similar approach is standard for NMR structural determination besides mapping experiments. The outlined approach works well in our experience and its accuracy was validated by x-ray structures in at least three cases[22-24], but it is qualitative and has some subjective terms. For these reasons, it is important to be conservative in the choice of residues to be included. Defining an incomplete epitope is better than including erroneous residues.

Broadening

The NMR peaks of antigen residues affected by antibody binding change position upon complex formation. Epitope signals, however, may also disappear upon binding. This is a consequence of "chemical exchange"[25], whose thorough description goes beyond the scope of this thesis. Briefly, an epitope residue alternating between free and bound state may

i) Generate two NMR signals, one corresponding to the free state and a different one for the bound state. This is called "slow exchange regime", with slow referring to the timescale of NMR experiments, and is usually expected for high affinity antibody complexes with slow dissociation rates. This situation is usually simple to analyze in NMR epitope mapping.

ii) In "fast exchange" epitope residues alternate between free and bound state before the NMR observation is completed; the result is a single signal in a position related to the average of the positions in the two states. This average position is different from the free state and can therefore be discerned in mapping experiments.

iii) The NMR signal of residues in "intermediate exchange" broadens as a consequence of their alternating between free and bound state. The effect can be so profound to make the signal of interface residues disappear below the noise. If certain antigen residues disappear upon complex formation, therefore, it might be tempting to conclude that they belong to the epitope. Although this may be the case, further considerations complicate the issue and recommend caution. The NMR signal, in fact, becomes broader not only because of chemical exchange but also due to increasing molecular weight. The effect,

significant in epitope mapping due to the large size of antibodies, is identical for all antigen residues: upon antibody binding the molecular weight of the observed species increases and all NMR peaks become broader regardless of them being part of the epitope, leading to partial or even complete loss of signal (compare the peak intensity of free and bound antigens in figure 2). To further complicate matters, signal broadening depends also on the structural flexibility of individual residues. Proteins have regions with different mobility. Beta strands in the protein core, for instance, are usually more rigid than protein loops. The combination of these factors may result, for example, in the signal of a strand residue disappearing upon complex formation due to broadening while a flexible loop residue remains visible. This depends strictly on local protein mobility and it would be erroneous to suggest that one of the two residues is part of the epitope. Overall, evaluating broadening effects in NMR epitope mapping is complicated and may easily lead to inaccurate epitope assignment. Even more so since it is often impossible to obtain a quantitative analysis of the broadening effects due to the poor spectral quality typical of antibody/antigen complexes. There are instances when including in the epitope a residue whose signal disappears upon complex formation might be appropriate, but they are rare in our experience. In the rare cases when we might use these considerations, we only do so towards the end of the iterative process for epitope discovery described above.

Does every affected residue belong to the epitope? Allosteric effects

Antigen residues whose NMR signal changes upon addition of an antibody are, indeed, affected by antibody binding. The NMR change, however, only reflects a difference in local chemical environment. This can derive from a direct contact with the antibody, meaning that a residue is part of the epitope, or from allosteric effects. If antibody binding to one site (epitope) provokes structural changes in a second site, then residues in this second site will have NMR chemical shift changes as well. Allosteric effects cannot be distinguished by NMR chemical shift perturbation experiments alone. Structural considerations and other experimental evidence should be sought to clarify the issue. Dedicated NMR experiments can help to differentiate direct antibody contact from allosteric effects but they have disadvantages over simple HSQC mapping both in term of sensitivity and sample preparation. In cross-saturation experiments, for instance, the antigen needs to be uniformly labelled with ^{15}N and ^{2}H while the antibody remains unlabeled[26]. The antibody protons of aliphatic sidechains are then saturated by an NMR pulse which affects them and the protons spatially close to them. The NH groups of the antigen close to the saturated antibody protons are thus affected and their NMR signal broadens, whereas long range allosteric effects have no impact. This allows identification of epitope residues directly in contact with the antibody without interference from allosteric effects. There are two main problems with cross-saturation experiments. If deuteration of the antigen is not complete, the remaining antigen protons will be saturated and antigen residues that are close to them, and not to the antibody, will be affected, leading to incorrect epitope determination. Furthermore, NMR cross-saturation

experiments are less sensitive than HSQC experiments, leading to lack of signal in the case of large antibody-antigen complexes. Finally, the conditions required to achieve complete deuteration are toxic for Escherichia coli or similar expression organisms, increasing the difficulty in obtaining sufficient amounts of the required protein sample. Deuterium exchange NMR studies are another strategy to avoid detecting allosteric effects. If a ^{15}N labelled protein is dissolved in D_2O, NH backbone protons exchange with the solvent and become ND. Since deuterium does not produce an NMR signal at the frequencies monitored in a ^{15}N-HSQC, their peaks disappear. Interface residues in an antibody/antigen complex, instead, may be buried and not accessible to the solvent; their NH groups, therefore, would not exchange with the solvent and remain detectable in a ^{15}N-HSQC. Comparison of the spectra of the free and complexed antigen dissolved in D_2O allow identification of residues that are shielded from solvent exchange only in the complex and are, therefore, likely to belong to the antigen/antibody interface or, in other words, be part of the epitope. Residues located in the protein core are easily identified because they are protected from solvent exchange both in the free and complexed antigen. The main drawback of this approach is that an antibody/antigen complex must be formed in solution, then ideally lyophilized and re-dissolved in D_2O. Not all antibody/antigen samples tolerate this harsh process. Extensive dialysis may be used in alternative to lyophilization[25].

Practical considerations

Antibodies are molecules of approximately 150kDa that include two identical antigen binding sites. Their large size is often detrimental to NMR investigation for reasons illustrated above. One option to reduce their size is to digest them to so-called Fab fragments; these are approximately 50kDa molecules with only one antigen binding site. Their production requires a straightforward enzymatic digestion[27] and purification step and commercial kits are available for this. Drawbacks of this procedure are the costs, both for reagents and due to the loss of material, and the fact that Fab fragment lack avidity effects that may be important for antibody function. Avidity indicates the ability of a single antibody molecule to bind the same antigen with both its binding sites at the same time, usually leading to increased affinity. However, in our experience the smaller molecular weight of Fab fragments offer a significant sensitivity advantage in NMR epitope mapping studies, so we prefer working on Fab rather than full antibodies whenever possible. The antibody size can be further reduced if it is produced as a single chain version of its variable region. This fragment (scFv) contains one antigen binding site and is approximately 25kDa[28]. scFv are not natural molecules and need to be engineered and produced in Escherichia coli or other systems. This is inconvenient (and often impossible due to refolding and aggregation problems) when they are derived from a full, natural antibody. However, artificial methods for antibody discovery and optimization like yeast display libraries[29] generate scFv molecules. The small molecular weight of scFv fragments often yields the best spectra in NMR epitope mapping experiments.

However

i) These molecules are prone to aggregation that may actually increase their apparent size to more than that of Fab fragments and

ii) Their production if often laborious and far from guaranteed. NMR epitope mapping experiments on full antibody, Fab fragment and scFv fragment are shown in figure 2.

As previously stated, we find it most advantageous to perform NMR epitope mapping using ^{15}NHSQC experiments. These experiments are very sensitive, which is important in large antibody/antigen complexes, are simple to analyze and yet provide a thorough residue-level description of the epitope. We typically run them for less than an hour on the free antigen and for 16/24 hours on the antibody/antigen complex, in order to increase the signal to noise ratio. We usually avoid other labelling options for simplicity of sample preparation. Finally, just as in any NMR experiment, buffer conditions, temperature etc need to be screened to avoid aggregation or other problems, and NMR parameters need to be optimized for sensitivity.

Conclusions

Solution NMR spectroscopy is ideally suited to the characterization of intermolecular interfaces, including antibody/antigen complexes. NMR chemical shift mapping provides a detailed description of the binding footprint of one molecule over another. Comparing the NMR spectrum of an antigen free and in complex with an antibody allows the determination of epitope

residues. The results are accurate and as detailed as those obtained from x-ray crystallography, as it was shown when NMR mapping of a TCR/pMHC complex (molecules very similar to antibody/antigen complexes) was validated by a later x-ray structure[24]. The NMR binding footprint of the TCR over the pMHC, equivalent to an antibody epitope, is in agreement with the contact map derived from the crystallographic structure (figure 6)

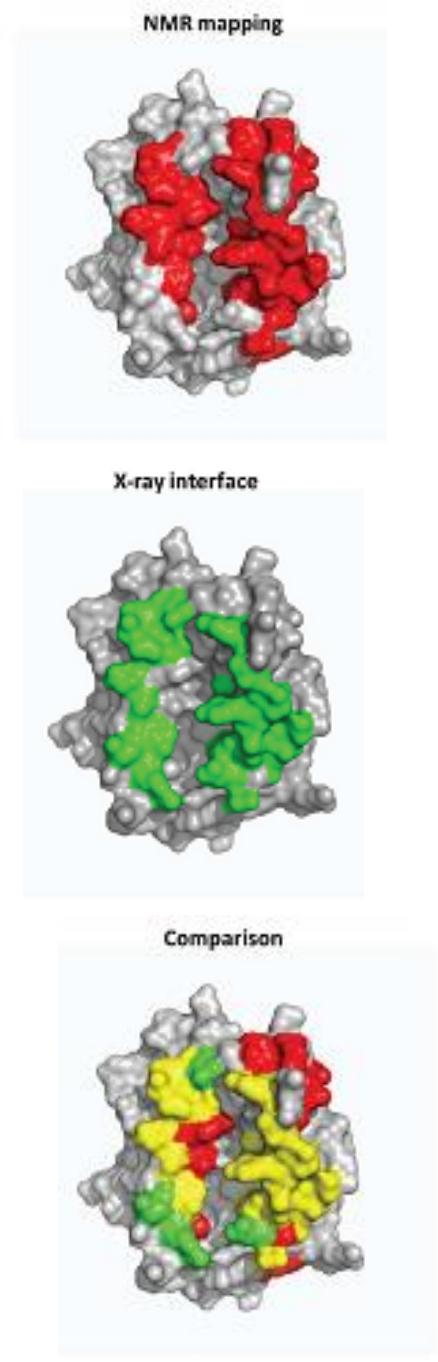

Figure 6: Comparison between an intermolecular interface determined by NMR mapping and x-ray crystallography (TCR/pMHC complex, which share many similarities with antibody/antigen complexes (Varani, Bankovich et al. 2007). The MHC structure is shown in grey, surface representation. MHC residues whose NMR signal is affected by TCR binding are shown in red on the top; MHC residues within 5A of the TCR in the x-ray structure are shown in green in the middle. Interface residues according to both NMR and x-ray are shown in yellow on the bottom. The rightmost structure shows good agreement between NMR and x-ray mapping. Some short range allosteric effect can be seen in the top right, red residues in the comparison picture.

In other works, the results of NMR epitope mapping were used to greatly increase the accuracy of computational docking, providing a three-dimensional atomic structure of the antibody/antigen complex. Roughly speaking, information on the NMR derived epitope is either used as a constraint to guide the structure calculation algorithm[30-32], possibly with other information that can help to orientate the components of a multi-molecular complex[33], or at the end of the computational simulation to filter out models that do not agree with the experimental epitope[34,35]. The resulting structures were sufficiently accurate to allow rational modifications of an antibody, altering its selectivity and increasing its viral neutralization properties by up to 40 folds[36]. Such rational engineering is often thought to require an x-ray structure. Although x-ray is obviously better at structure determination, NMR mapping offers comparable results in a fraction of the time for epitope

characterization. Once NMR assignments are available for the antigen several different antibodies can be rapidly mapped, typically requiring a couple of days per antibody. NMR epitope mapping also has a clear advantage over peptide mapping and site directed mutagenesis, since it provides a complete description of the antigen residues contacted by the antibody which is very difficult to obtain by these other techniques. The disadvantages of NMR epitope mapping include the fact that the antigen needs to be labelled with NMR active nuclei (typically ^{15}N). It is thus necessary to express the antigen in Escherichia coli (more seldom yeast), which is not always possible. Furthermore, large antigens present a challenge for NMR, due to both spectral overlap and signal broadening at increased molecular weight leading to loss of sensitivity. Selective labelling of specific antigen residues, advanced acquisition techniques like TROSY and the use of ^{15}N-HSQCs, probably the simplest and most sensitive multidimensional NMR experiment, somehow alleviate these limitations, which nonetheless remain significant in some cases. Finally, NMR assignment of the antigen, i.e. knowing which NMR signal corresponds to which antigen residue, is necessary. Assignment can be a trivial process of a few days or a difficult task requiring several months and complex strategies for selective labelling. If assignment is not available, NMR epitope mapping can still be useful to map the relative position of multiple antibodies. In a sort of residue level cross-competition experiment, it is possible to see if the same peaks (corresponding to antigen residues) are affected by binding of different antibodies. NMR epitope mapping is a rapid and accurate tool for the characterization of antibody/antigen complexes. So far there are relatively few examples of using high resolution NMR

spectroscopy to determine antibody epitopes and most involve small peptides rather than full protein antigens[35-42] but we are strongly convinced that NMR epitope mapping is a valuable tool for the detailed structural characterization of antibody epitopes.

Section II

Computational

Docking

Abstract

Molecular docking is a popular and useful tool in the drug design; Ligand binding is the key step in enzymatic reactions and, thus, for their inhibition. Therefore, a detailed understanding of interactions between small molecules and proteins may form the basis for a rational drug design strategy[43-46], as well as a better understanding of interaction between antibody and antigen through specific atomic interactions between the antibody and the region of the antigen (Ag) that it recognizes (epitope) 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.

Introduction

The best way to study atomic interaction is to obtain the three-dimensional structure of antibody-antigen complexes. Traditionally, this is achieved by X-ray crystallography technique, an often long and laborious process with high failure rate. Thanks to advances in algorithms and processing power, now we are able to use computational techniques for the structural characterization of intermolecular complexes.

Computational docking is the process of predicting the structure of a complex starting from the separated structures of its individual components[47], it is emerging as a fast and affordable technique for the structural characterization of antibody-antigen complexes and protein-protein complexes. 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.

Computational Docking

The ability of the docking prediction is to sample many state of a complex, determining the free energy, called score, for each of them, so the predicted structures having lower free energy are the most likely realistic three-dimensional structure for the complex, considering the hypothesis that the structure of a complex is the lowest free energy state accessible to the system[48].

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. Usually the bigger structure of the 2 partners is fixed in the space and the smaller one is moved in the direction of the previous one with small movements, each new movement create a new energy state (score) for the computational complex. This is an iterative process continuing up to find the lowest possible score for the computational complex. This iterative process generate thousands of computational complex structures each one with a specific score.

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.

(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[49].

(iii) Modeled, if it has been predicted by homology modeling and/or other computational techniques like *ab initio* predictions or molecular dynamics.

Antibody modeling

In order to predict an antibody structure is important remind that the structure of the antibodies is governed by very precise rules, especially with respect to their antigen binding loops, called complementary determining region (CDR), and on the role of individual amino acid residues; in effect these information are critical for the antibody design and provides the rational basis for any antibody structure prediction method.

The Antibodies are globular plasma glycoproteins of 150 kDa molecular weight. The monomer is a "Y"-shaped molecule that consists of four polypeptide chains: two identical heavy chains and two identical light chains connected by disulfide bonds. Each chain is formed by structural domains, called immunoglobulin domains, that contain about 100 amino acids and have a characteristic fold in which two beta strand create a "sandwich" shape, kept together by disulfide bond and conserved electrostatic interactions[50,51].

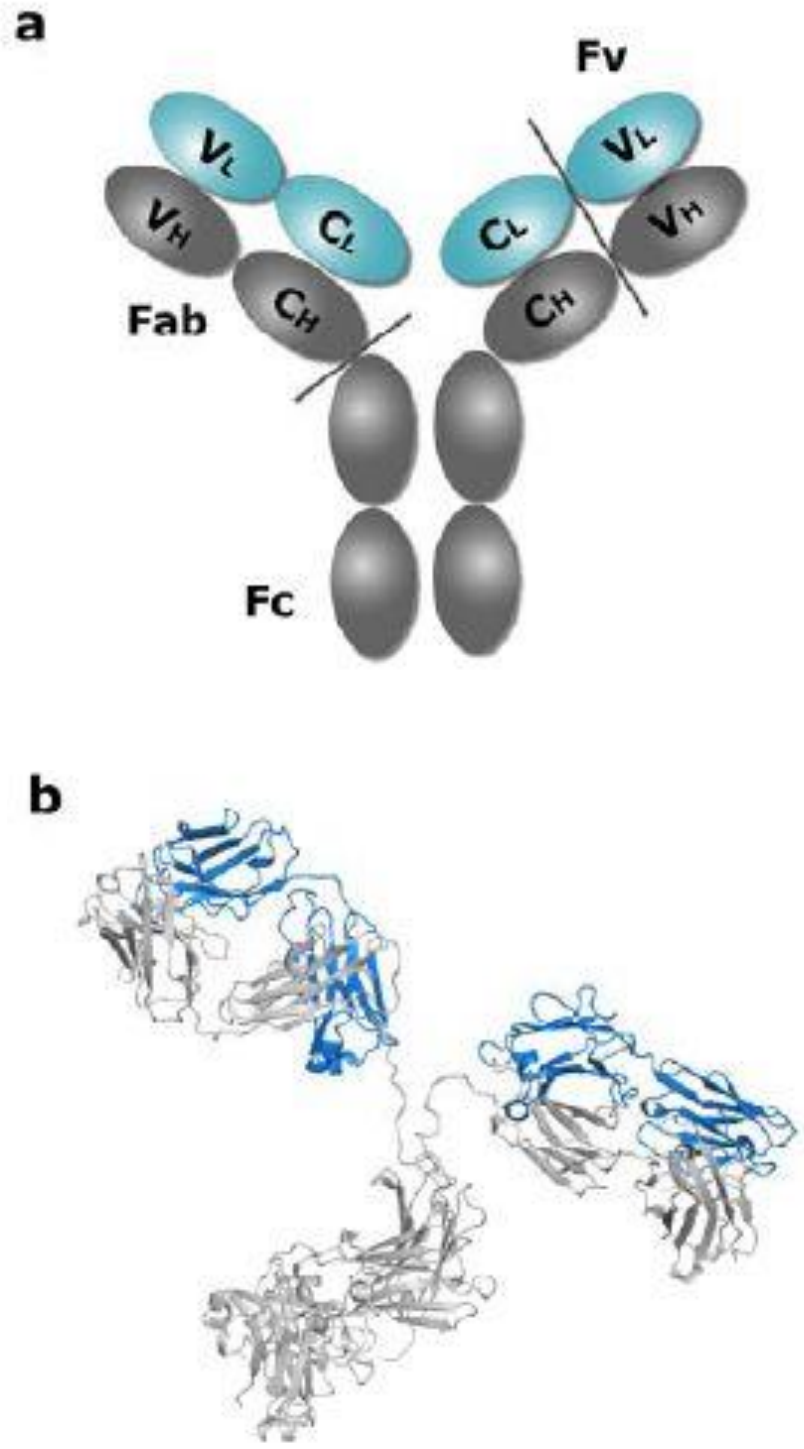

Figure 7: Schematic (a) and cartoon (b) representation of a full antibody structure. Antigens bind to the tip of the V_H and V_L domains.

The arms of the Y is called Fab (fragment antigen binding) and can bind two antigens, it is composed of one constant and one variable domain from each heavy and light chain of the antibody. The variable domain is also referred to as the F_v region and contains the CDR loops, three loops on each of the light (V_L) and heavy (V_H) chains are responsible for binding to the antigen; these loops are also referred to as H1,H2, H3 and L1, L2, L3.

The base of the Y is called the Fc (Fragment crystallizable) domain, and is composed of two heavy chains and ensures that each antibody generates an appropriate immune response for a given antigen by binding to a specific class of Fc receptors and other immune molecules such as complement proteins[52].

The canonical structure of the antibodies make them suitable candidates for modeling. The most variable part formed by CDR loops can assume just a small repertoire of main-chain conformations, called "canonical structures"[53-55].

These conformations are characterized by the length of the loops and by the presence of key residues at specific positions in the antibody sequence.

The canonical structure method is used for the prediction of the antibody structure for the most of the web-server, PIGS (Prediction of Immuno Globulin Structure) or WAM (Web Antibody Modeling server)[56,57].

Antibody Modeling Based on Canonical Structures

In the canonical method, the sequence of each variable domain V_L and V_H of an unknown structure is aligned with the variable domain sequence of all the immunoglobulin of known structure, using standard database searching and multiple sequence alignment programs (BLAST, Clustalw)[58,59].

The backbone structure of the framework is modelled using as template known structure having higher sequence identity, because higher is the residues identity in the core of the two proteins more similar is the conformation and also the quality of the model[60].

The prediction of the CDR loop, critical for the antigen binding is also predicted using templates having the same loop conformation and sequence identity and combining the templates in order to obtain the most accurate modeling. Different combinations of templates can be used as showed 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 modelled 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[60-62]. 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.

Afterwards the modeling of the side chains happen. At the sites where the residues are the same, the conformation of the parent structure is retained. At the sites where the residues are different, the side chain conformation is copied from antibodies with high sequence similarity or imported from standard rotamer libraries[63].

Finally the model is refined by few cycles of energy minimization to improve the stereochemistry, especially in those regions where segments of structure

coming from different immunoglobulin have been joined and not to refine the models significantly.

Rosetta Antibody Modeling

Rosetta Antibody[64] is a homology modeling program to predict antibody F_v structures. It uses a simple energy function to simultaneously optimize the CDR loop backbone dihedral angles, the relative orientation of the light and heavy chains and the side chain conformations. The program can be downloaded and run on local computers or modeling requests can be submitted to a web server[65]. Rosetta Antibody first identifies the antibody templates with highest sequence identity for each framework and CDR loops; the loop templates are then grafted onto the framework and the full F_v is assembled. This crude model is used as input for a second stage: a multi-start, Monte-Carlo-plus-minimization algorithm that generates two thousand candidate structures. H3 loop conformations are generated by assembling small peptide fragments[66] and sidechains are finally optimized via rotamer packing and energy minimization[67]. The CDR backbone torsion angles and relative orientation of light and heavy framework are also perturbed and minimized with a pseudo-energy function that includes van der Waals energy, orientation-dependent hydrogen bonding[68], implicit Gaussian salvation[69], side chain rotamer propensities[70] and a low-weighted distance-dependent dielectric electrostatic energy[71]. In the end, a scoring function is used to discriminate the 10 best antibody models that are offered as standard result.

The Docking Calculation

Starting from the chosen or generated structures for the antibody and antigen, the molecules are then brought together by the preferred algorithm. The assumption is that the biological structure having the lowest energy is also the correct one because is energetically favored. The scoring function, associated to each model coming out of the computational simulation, try to simulate an energy by counting for biophysical consideration such as hydrophobic and electrostatic interaction, salt bridges, hydrogen bonds, but also statistical and empirical consideration such as the degree of conserved residues to the interface.

When searching for the correct binding orientation, the two molecules are allowed to move and the score is assessed after each step. A Monte-Carlo minimization protocol retain conformation with higher energy in an attempt to overcome local energy minima that don't correspond to the global minimum. The movement is stopped after a predefined number of steps or when the score does not improve further. The conformation parameter are changed between each step vary in different docking algorithms, which may be divided in three general classes as described below:

(i) Only the relative position of the docking partner is changed;

(ii) The relative position and the sidechain conformations are changed;

(iii) The backbone conformation is altered in addition to the above.

In the simplest case, is called rigid body docking because the conformation of the starting structures is not altered at all during the docking process and the scoring function only needs to account for the intermolecular interactions[72,73].

RosettaDock has a first rigid body phase in which sidechains are removed, but in a second phase they are re-introduced and their orientation is optimized[74,75]. Since the sidechain conformation is dictated mainly by a limited number of allowed torsion angles, the task can be completed with reasonable success and limited computational requirements[76]. Accurately simulating the backbone movements that often happen upon formation of biological complexes, instead, remains a daunting task for docking, which has a very high failure rate when molecules undergo significant conformational changes upon binding. It is conceivable that in vivo antibodies adapt to and are selected against existing antigen conformations[77], thus it might be tempting to believe that antigens should not experience drastic changes upon antibody binding. Rigid body docking might be best in this case but first of all it is doubtful that proteins are not subjected to any conformational motion in solution, not even at the sidechain level, and furthermore there are examples in which antibodies provoke relatively large allosteric effects on the antigen.

The issue is slightly different for the antibody, instead: since antibody modeling uses bound conformations as templates, the conformational rearrangements experienced by the antibody upon binding can be ignored. It should be noted, however, that the canonical structures used for antibody modeling describe the backbone but not the side-chain conformations, which are probably best explored during the docking run.

In conclusion, if one believes that the antibody model is accurate and that antigen binding loops are relatively rigid, then it should not be necessary to sample antibody backbone flexibility in the docking run. This assumption appears reasonable for the 5 CDR loops following canonical structural rules

but it might fail for the H3 loop, the third CDR loop of the heavy chain, which may be slightly inaccurate and/or might indeed be flexible in the biological context. Conversely, docking methods that vary the CDR loops conformation might introduce deviations from the canonical structure and decrease the accuracy. Although it is impossible to draw general rules, using rigid body approaches for the backbone but sampling different sidechain conformations might be a reasonable compromise. It might also prove useful to allow backbone movement for the H3 loop (and others when they do not follow canonical structures) while allowing only sidechain optimization of the remaining antigen binding loops. This behavior can also be approximated by generating multiple antibody models, presumably differing mainly in the H3 conformation, and using all of them as starting structures to be docked without backbone optimization[78], either as an ensemble or serially.

Section III

Structural characterization of the interaction between CXCL12 and HMGB1

Abstract

Each biomolecule in a living organism needs to adopt a specific three-dimensional conformation to function properly. Function itself is usually achieved by specific interactions between bimolecular units. Structural knowledge at atomic level of biomolecules and their interaction is important to understand mechanisms leading to biological response and to develop strategies to interfere with them when necessary.

The present study uses a combination of solution NMR mapping, molecular biology and computational docking to investigate the structure of the chemokine CXCL12 in complex with the chromatin-associated protein HMGB1.

CXCL12 binds to the chemokine receptor CXCR4 and plays an essential and unique role in homeostatic regulation of leukocyte traffic. High-mobility group box 1 (HMGB1) protein mediates activation of immune responses including chemotaxis and cytokine release.

HMGB1 exerts its chemotactic activity by forming a hetero-complex with CXCL12, which acts exclusively through its receptor CXCR4 and not through other HMGB1 receptors.

This study proves that the two proteins indeed interact with each other, resulting in an increased CXCL12 cellular activity and maps the regions of the two proteins necessary for either interaction or activity.

Introduction

Chemokines and their receptors

Structure

Chemokines are a group of small proteins ranging from 67 to 127 amino acids (8-12 kDa), their sequence is very variable but all share very similar tertiary structure. The chemokine family has a characteristic cysteine motif consisting, most of the times, of four cysteine residues with two disulfide bonds linking the first cysteine with the third, and the second with the fourth, crucial for their three-dimensional fold[79]. NMR and X-ray structures revealed the conserved tertiary structure of chemokines, which includes a disordered N-terminus of 6-10 amino acids that functions as a key signaling domain in all chemokines. The N-terminus is followed by a loop ending in a 3_{10} helix, connected with three antiparallel β-sheet strands and a C-terminal helix. The above mentioned disulfide bonds stabilize the overall topology.

Many chemokines form dimers or higher order oligomers in solution or upon binding to glycosaminoglycans (GAGs). In the so-called CXC chemokines residues in the first β-strand from one subunit interact with the same strand from a second subunit, forming a single, extended six-stranded sheet. The overall dimer topology is a β-sheet platform topped by two α-helices[80-87].

Cellular Function and chemokine receptors

Chemokines play a key role in the migration, homing and retention of immune cells. Chemokines bind to specific seven-transmembrane G-protein coupled receptors (GPCRs) expressed on the surfaces of their target cells. Binding triggers signal transduction events culminating in a multitude of cellular responses, such as flux of intracellular calcium (Ca^{2+}) ions, cytoskeletal rearrangements or activation of cell adhesion molecules. Cell migration is also controlled by expression of different chemokine receptors on the cell surface, which recognize specific signals allowing movement of a cell towards an inflammation site[88].

Some chemokine receptors bind only specific chemokines, such as the CXCR4 (receptor) - CXCL12 (chemokine) pair; other receptors can bind several chemokines with different affinity. Peptides derived from the N-terminal domains of chemokine receptors bind specifically to their respective chemokine ligands[85,89].

More than 20 chemokine receptors and 40 chemokines have been identified so far (table 1)[90].

Receptor	Ligands	Expression	Function
CCR1	CCL3,5,7,8,13, 14,15,16, 23	T cells, monocytes, neutrophils	Inflammatory recruitment
CCR2	CCL2, 7, 8,13	T cells, monocytes, macrophages	Inflammatory recruitment
CCR3	CCL5, 7 ,8, 11, 13, 15,24, 26	Eosinophilis, basophilis, mast cells, Th2 cells	Inflammatory recruitment
CCR4	CCL17, 22	Th2 cells	Skin homing
CCR5	CCL3, 4, 5, 8, 14	Dendridic cells, Th1 cells, monocytes, $CD34^+$ progenitors	Inflammatory recruitment, HIV entry
CCR6	CCL20	Th17 cells	
CCR7	CCL19, 21	Lymphocytes, mature dendritc cells	Lymph node/T cell zone homing
CCR8	CCL1, 4, 17	Thymus Th2 cells	Inflammatory recruitment
CCR9	CCL25	T cells	Gut, thymic, t cell progenitor homing
CCR10	CCL26, 27, 28	T cells	Skin homing
CXCR1	CXCL8	Lymphocytes , monocytes, macrophages	Inflammatory recruitment
CXCR2	CXCL1, 2, 3, 5, 6, 7, 8	Neutrophils, monocytes, macrophages	Inflammatory recruitment
CXCR3	CXCL9, 10, 11	Activated T cells	Inflammatory recruitment
CXCR4	CXCL12	Hematopoietic cells, endothelial cells, neurons, microglia, astrocytes	Embryogenesis, hematopoiesis, bone marrow homing, inflammation, HIV entry
CXCR5	CXCL13	B cells, follicular B helper T cells	B follicle recruitment
CXCR6	CXCL16	Activated T cells	Inflammatory recruitment
XCR1	XCL1, 2	Lymphocytes, neutrophils	
CX_3CR1	CX_3CL1	T cells, NK cells	Cell adhesion, inflammatory recruitment

Table 1: Chemokine receptor ligand, expression and function.

Functional criteria divide chemokines in inflammatory or homeostatic classes.

Inflammatory chemokines are expressed by leukocytes and other cells upon activation and have role in both innate and adaptive immunity in response to infection and tissue damage. Expression of inducible chemokines is triggered

by inflammatory mediators such as tumor necrosis factor, interferon-γ, microbial products or trauma. Their expression is temporary and stops after resolution of the situation.

Homeostatic chemokines are constitutively expressed and coordinate the basal level of cell migration needed for proper function of the immune system. They navigate leucocytes during haematopoiesis in the bone marrow and thymus, during initiation of adaptive immune responses in the spleen and lymph nodes, and in immune surveillance of healthy peripheral tissues. Recent findings indicate that several chemokines cannot be assigned unambiguously to either of the two functional categories and, therefore, may be referred to as "dual function" chemokines[90].

CXCL12

Structure

CXCL12 exists in a monomer-dimer equilibrium that shifts towards dimer in presence of various binding partners, including the extracellular domain of its receptor CXCR4. The monomeric conformation is favored in acidic conditions.

CXCL12 structure includes a flexible N-terminus connected by an extended N-loop and a turn of 3_{10} helix to a three-stranded β-sheet and a C-terminal α-helix[91] (Figure 8). The flexibility of the α-helix gives CXCL12 a very dynamic structure sensitive to solution changes and allowing the switch between monomeric and dimeric forms.

Residues in the flexible N-terminus of CXCL12 are required for CXCR4 binding, which promotes dimerization of the former, and activation[92]. CXCL12 binds to the extracellular N-terminal domain of CXCR4[93], which must be post-translationally modified by sulfation at three tyrosine residues (Tyr7, Tyr12, and Tyr21)[94,95].

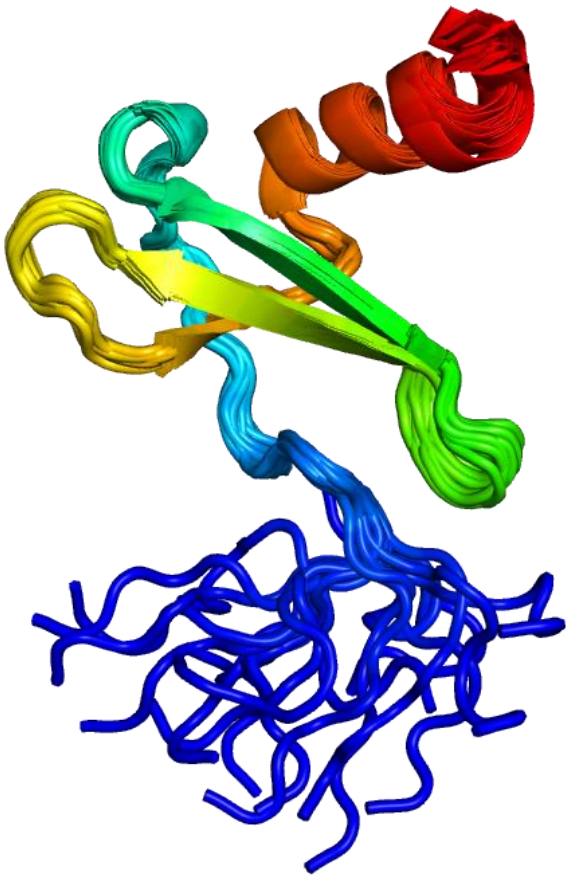

Figure 8: NMR structure of monomeric CXCL12 (PDB ID 2KEE).

Cellular Function

CXCL12 is a potent chemo-attractant for lymphocytes, monocytes and circulating neutrophils[96]. CXCL12 signaling via CXCR4 regulates the development of T and B lymphocytes and contributes to the survival of mature lymphocytes and to the generation of memory T cells[97]; it also plays a role in the homing of hematopoietic progenitors cells to the bone marrow microenvironment[98]. Moreover CXCL12 and CXCR4 are involved in the embryonic development of different peripheral tissues including blood vessels, muscles and primordial germ cells[99-101]. Several studies indicate that the CXCL12/CXCR4 interaction enhances the inflammatory infiltration of lymphocytes in different pathogenic processes[102].

The chemokine receptor CXCR4 was the first identified co-receptor for HIV virus that infects $CD4^+$ T-lymphocytes[103]. In addition CXCR4 is constitutively expressed in human brain by neurons and microglia, the principal cell type infected by HIV-1 in the CNS. The up-regulation of CXCR4 expression has been reported in macrophages, microglia and neurons in AIDS patients[104-106]. HIV-1 infected microglia or macrophages regulate astrocyte CXCL12 production through IL-1β[107], providing the ligand for up-regulated CXCR4. The abnormal expression of CXCL12 and CXCR4 in HIV is consistent with the possibility that their interaction might contribute to the pathogenesis of HIV-1-associated encephalopathy.

By contrast, CXCL12 can inhibit HIV-1 infection in cells of the immune system by inducing receptor internalization, thus effectively diminishing the number of CXCR4 molecules on the cell surface. Furthermore CXCL12, similarly to small molecule CXCR4 antagonists (*e.g.* AMD3100) or CXCR4 specific antibodies (*e.g.* 12G5), can inhibit HIV-1 infection *in vitro* by competing for the binding of the viral envelope protein gp120 to the receptor[108].

HMGB1

Structure

HMGB1 is a 215 amino acids protein of ~30 kDa composed of three domains: two positively charged domains (BoxA and BoxB) connected by a flexible linker and a negatively charged carboxyl terminus (the acidic tail). HMGB1 adopts a closed, dynamic but compact conformation, as shown by nuclear magnetic resonance spectroscopy and small angle X-ray scattering[109,110]. The long acidic tail of HMGB1 is flexible and lacks a secondary structure and interacts with specific residues in the rest of the protein[111]. BoxA has three α-helices folded into an L-shape, whereas BoxB has four α-helices folded in L-shape[112] (figure 9).

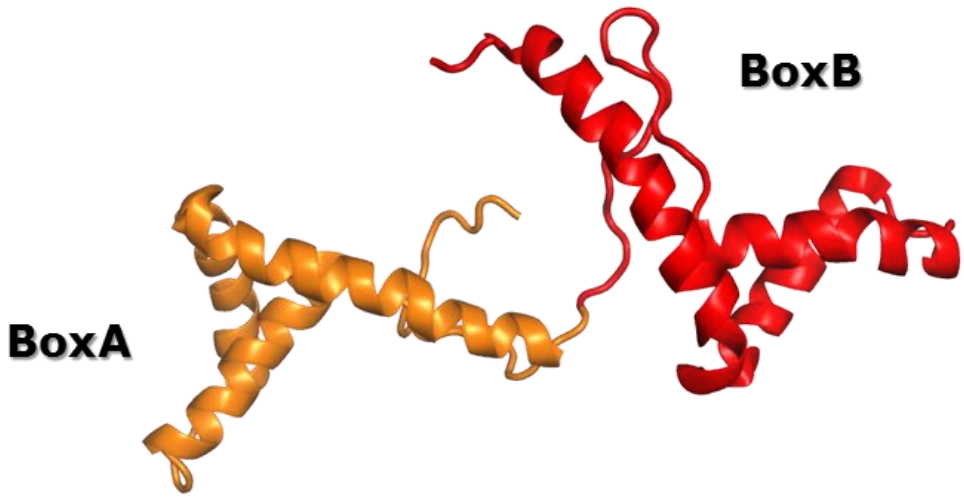

Figure 9: Solution structure of the tandem HMG box domain from Human High mobility group protein B1 (PDB ID 2YRQ).

Cellular Function

In the cellular nucleus HMGB1 induces DNA bending by local DNA distortion and facilitates the binding of several regulatory protein complexes to DNA. Through BoxA and BoxB, HMGB1 enhances transcriptional activation and regulates several families of DNA-binding proteins such as the p53–p73 transcriptional complexes[113,114]. When not bound to DNA, the long acidic tail of HMGB1 interacts with basic stretches in BoxA and BoxB, preventing other interactions[111].

HMGB1 is expressed early during development and it is also widely expressed in the adult organism, particularly in immune cells of the thymus and in circulating monocytes[115]. HMGB1 is critical for normal development; indeed it was shown that knockout mice suffers from multiple organ failure, small size, absence of fat and impairment in bone formation. hmgb1-/- mice die at most 24 hours after birth due to severe hypoglycaemia[116,117].

HMGB1 interaction with chemokines

In addition to the role played in the nucleus, HMGB1 is also found also in cell cytoplasm[118] and extracellular space[119]. HMGB1 can be localized to the cell periphery and secreted from monocytes, macrophages, dendritic cells and fibroblasts. The active secretion of HMGB1 is mediated by acetylation in the two major clusters of lysines that act as nuclear localization signals (NLSs). Acetylation prevents the interaction of HMGB1 with the nuclear-importer protein complex, blocking re-entry into the nucleus[120].

Alternative cytosolic translocation has been observed in neutrophils in which HMGB1 is mono-methylated at lysine 42. Methylation supposedly changes the HMGB1 conformation and weakens its DNA binding activity, causing it to become largely distributed in the cytoplasm by passive diffusion out of the nucleus[121].

Finally, HMGB1 can be actively secreted outside cells through stimuli by pathogenic LPS and pro-inflammatory cytokines such as IL-1 and TNF. It can also be passively secreted from necrotic cells diffusing out into the extracellular space[122,123]. The increase of extracellular HMGB1 concentration is mainly related to pathological conditions such as acute and chronic inflammation, cancer and trauma[124,125].

Goal of the project:

HMGB1 interacts with CXCL12 and increases its cellular activity?

It was suggested that HMGB1 can directly and specifically interact with CXCL12, increasing its ability to bind to CXCR4 and consequently increasing its biological activity. The observation derived mainly by cell migration experiments conducted by our collaborators (Uguccioni group, IRB) and thus not included in this thesis. Briefly, CXCL12 can induce migration of cells expressing the CXCR4 receptor. It was observed that the migration increased in the presence of HGMB1 or, more exactly, that in the presence of HMGB1 a smaller amount of CXCL12 is required to achieve comparable migration. The effect is evident only in the presence of the full HMGB1 protein and not with the isolated BoxA or BoxB. Co-immunoprecipitation and other biochemical approaches could not provide any conclusive evidence for interaction.

During my thesis we set forth to 1) demonstrate that HMGB1 can directly interact with CXCL12 and enhance its cellular activity and 2) provide a structural description of the HMGB1/CXCL12 complex.

Results

We used solution NMR spectroscopy to confirm and characterize the interaction between CXCL12 and HMGB1, mainly through chemical shift mapping: a highly sensitive tool for characterizing intermolecular interactions, mapping binding sites and detecting interacting residues.

NMR is particularly sensitive to interaction even between low-affinity binding partners; simple and rapid ^{15}N-HSQC experiments (see below) can unequivocally confirm interaction at the residue level.

Solution NMR spectroscopy

Solution NMR spectroscopy is ideally suited to the atomic level characterization of intermolecular interfaces. NMR mapping can be used to rapidly and accurately determine protein/protein interacting interface. Basically, the differences in the NMR signal of a protein free or bound to a partner will identify the residues involved in the binding and consequently the binding surface.

The NMR signal is exquisitely sensitive to the local chemical environment. When an intermolecular complex forms, the chemical environment of interface atoms changes as they get close to the atoms of the interacting partner and so does their NMR signal. By comparing the NMR spectrum before and after complex formation, interface atoms can be individuated due to changes in their NMR signal (figure 10).

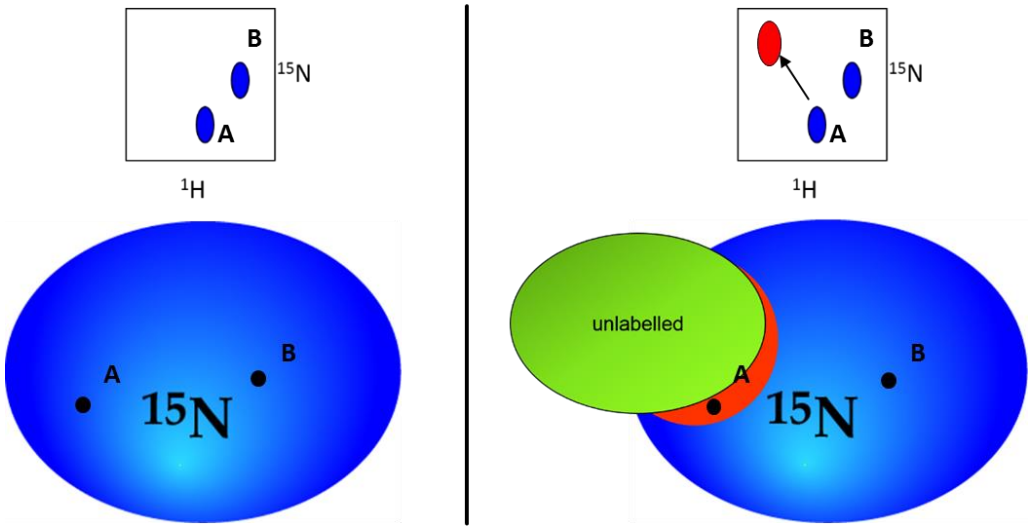

Figure 10: schematic representation of an NMR mapping experiment. Residues A and B of a 15N-labeled protein (left) generate a single peak in a ^{15}N-HSQC experiment, schematically shown at the top. Upon binding of a second molecule (right), interface residues experience a different chemical environment and their NMR signal changes, as shown for residue A. The interacting protein is unlabeled and its signals do not appear in ^{15}N-HSQCs. Interface residues whose signal is affected by binding can be identified by comparing the NMR spectrum of the protein free and in complex.

In a ^{15}N-Heteronuclear Single Quantum Coherence (HSQC) experiment[4] the backbone NH group of each protein residue (except prolines) generates a single NMR signal. The position of these signals is so sensitive to the protein conformation that ^{15}N-HSQC spectra are often referred to as protein fingerprints. Protein labelling with ^{15}N is required because the naturally occurring nitrogen isotope, ^{14}N, is not NMR active.

The idea behind NMR epitope mapping, schematically illustrated in Figure 3 and 4, is to (a) record a spectrum of the free, labeled protein, (b) add unlabeled

partner protein and record an HSQC of the complex. Because the partner protein is not labeled, its residues do not produce an NMR signal and thus do not appear in the spectrum. (c) Overlay and compare the spectra of the labelled protein free and in complex with the partner protein. If a signal changes position, it means that the chemical environment of the residue generating it has changed; in other words, the residue is affected by protein binding and is likely to belong to the interacting interface. The list of all protein residues whose signal changes upon partner protein addition provides a full description of the binding surface.

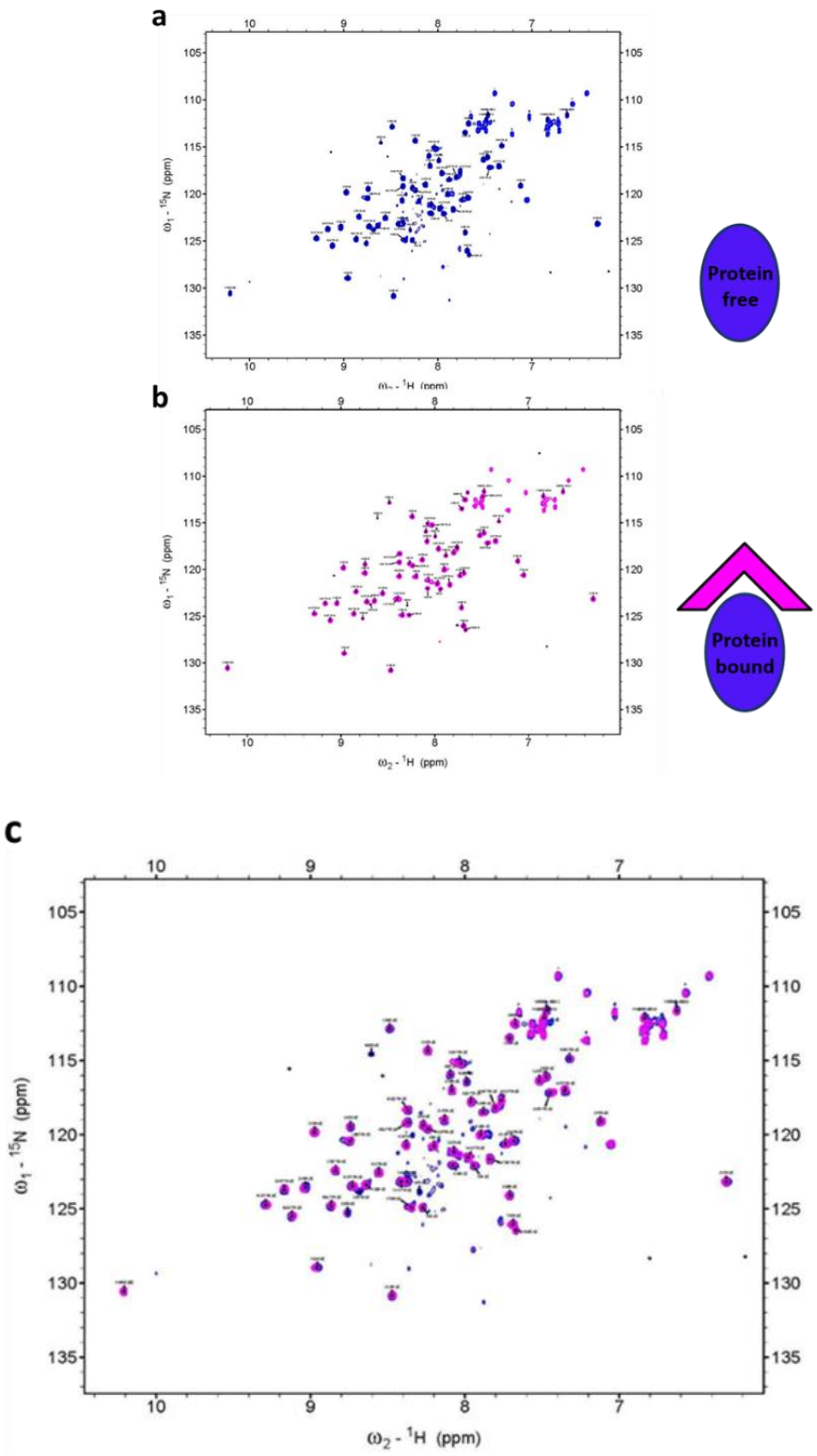

Figure 11: Schematic view of NMR epitope mapping experiment. a) HSQC spectrum of labelled protein free in solution, b) HSQC spectrum of labelled protein in presence of unlabelled partner protein, c) overlay of the two spectra.

The NMR assignment of the free protein of interest, i.e. knowing which residue generate a given NMR peak, is essential for NMR mapping. Assignment of the same protein in the complex is not strictly required. If NMR changes upon complex formation are minimal then assignment can be estimated by simple comparison with the free protein (figure 11). When changes are small it is usually trivial to pair each peak in the complex to its free counterpart. The larger and more numerous the changes, the less reliable this approach becomes. If a protein changes its structure upon complex formation, for instance, the NMR signals will change as well and it may be impossible to assign it by comparison to the free protein. If complex assignment is only available by comparison with the free spectra and ambiguities are present, one possibility is to consider the 'minimum shift difference'[19]. The amount of chemical shift change is defined as the difference between a given, assigned, peak in the free spectrum and its closest peak in the bound state. This guarantees that the entity of the chemical shift change is equal or larger to the one set in the analysis, but it is not overestimated.

NMR protein assignment requires the use of several aptly designed experiments and can require from a few days to years according to protein size and sample behavior. Since NMR assignments for both CXCL12 and HMGB1 are publicly available[91,112], we did not need to perform any such experiment during my thesis.

Can CXCL12 interact with the HMGB1 sub-units, BoxA and BoxB?

As a first step we aimed to ascertain whether CXCL12 can interact with either of the two HMGB1 structural domains, BoxA and BoxB. NMR signal intensity decreases and increasing molecular weight and spectral analysis becomes more and more complicated. Investigating the binding of the relatively small HMGB1 subunits allowed us to work in easier to deal with NMR conditions.

CXCL12, BoxA, BoxB and subsequently full HMGB1 were produced in Escherichia coli grown in M9 minimal medium either enriched or not with ^{15}N and purified according to standard conditions (see methods).

NMR ^{15}N-HSQC spectra of free ^{15}N-labeled CXCL12 were initially recorded. either BoxA or BoxB were subsequently added and ^{15}N-HSQC of the complex were recorded. The spectra of free and bound CXCL12 were compared and residues whose signal changed upon complex formation were identified. These residues are affected by binding and either part of the intermolecular interface or subjected to allosteric effects.

Complex between CXCL12 and BoxB

The NMR signal of 38 out of 68 CXCL12 residues shifted upon addition of an equimolar amount of BoxB, proving that the two molecules interact with each other.

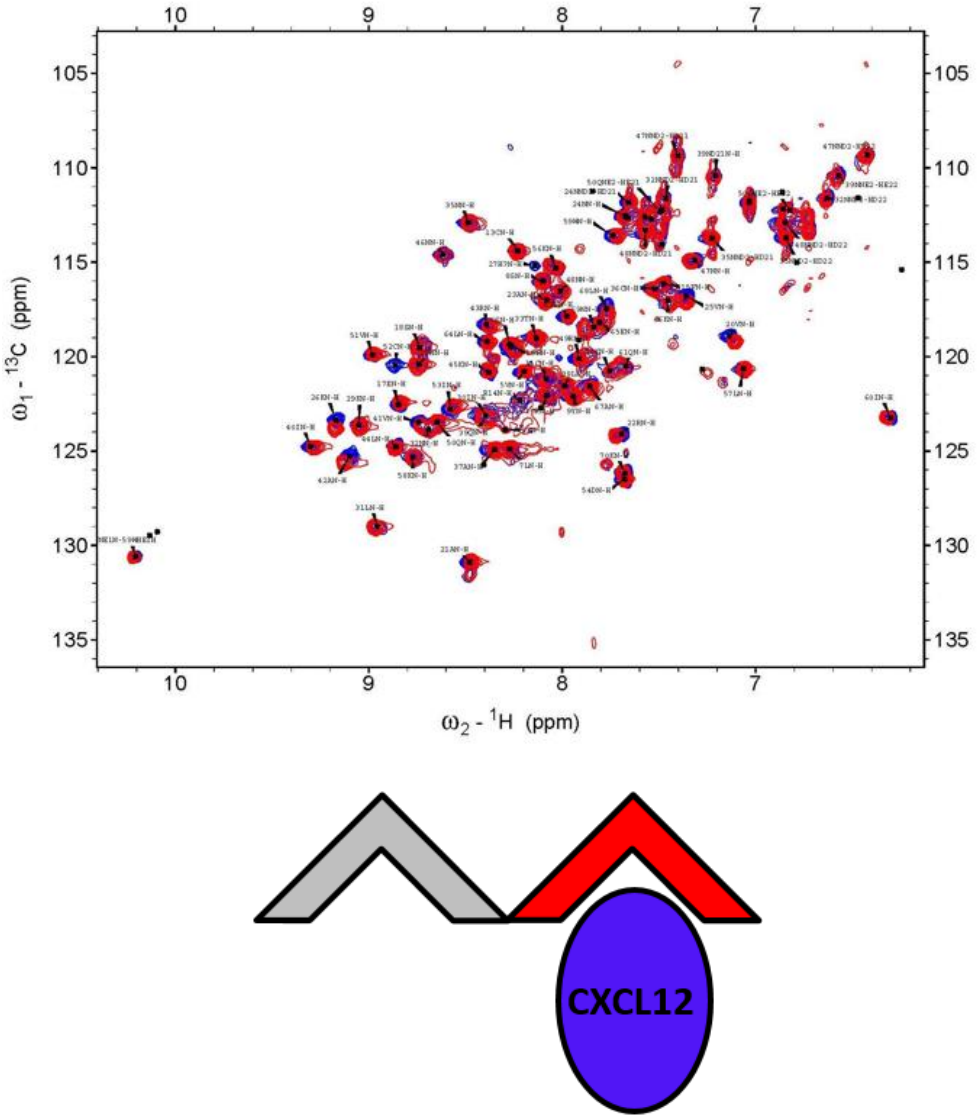

Figure 12: overlay of ^{15}N-HSQC spectra CXCL12 free in solution (blue) and in presence of 1:1 ratio of BoxB (red).

Residues whose NMR signal changes position upon complex formation were identified by visual comparison of free and bound spectra of CXCL12 (figure 12). A quantitative analysis was also performed: the "Chemical Shift Perturbation" index (CSP), i.e. the amount of chemical shift change upon complex formation, was calculated for each residue according to the formula $\sqrt{(\Delta N \cdot 0.2)^2 + \Delta H^2}$, where ΔN and ΔH are the chemical shift difference between free and bound spectrum in the nitrogen and proton dimension (figure 13).

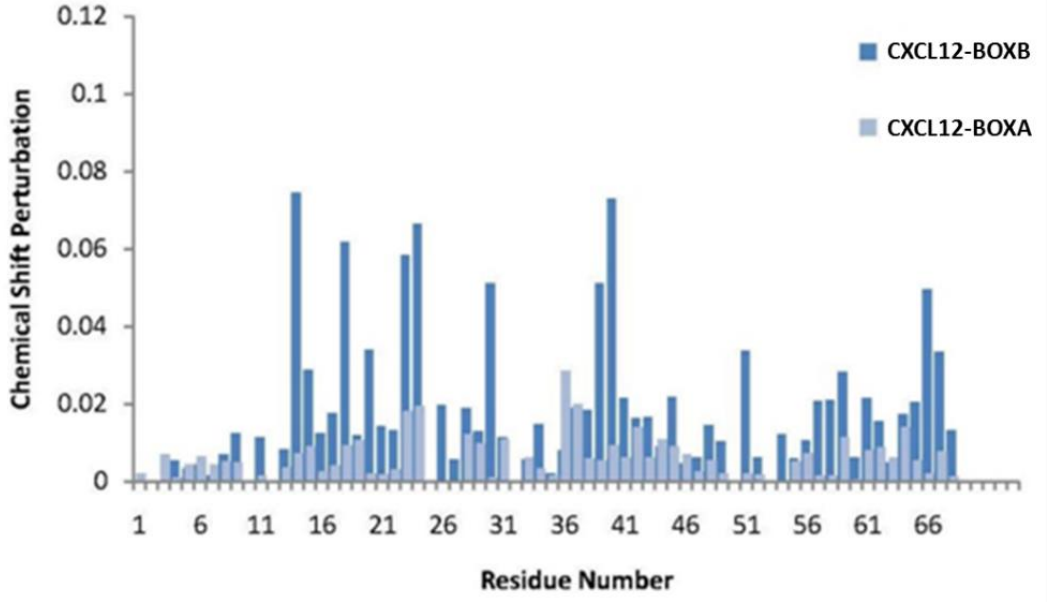

Figure 13: Chemical shift perturbation index (y axis) plotted against the CXCL12 sequence (residue number shown in the x axis). CSP for the CXCL12/BoxB complex are in dark blue and the CXCL12/BoxA complex in light blue.

The majority of the changes in CXCL12 involve residues 13–25, 40–45, and the C terminus, whereas the N terminus of CXCL12, in particular the flexible first 8 residues known to be required to trigger the CXCR4 receptor cellular

activity, are clearly not affected by BoxB binding. Mapping the residues with larger CSP on the CXCL12 structure defines a contiguous interacting region compatible with a footprint of BoxB dimensions (figure 14).

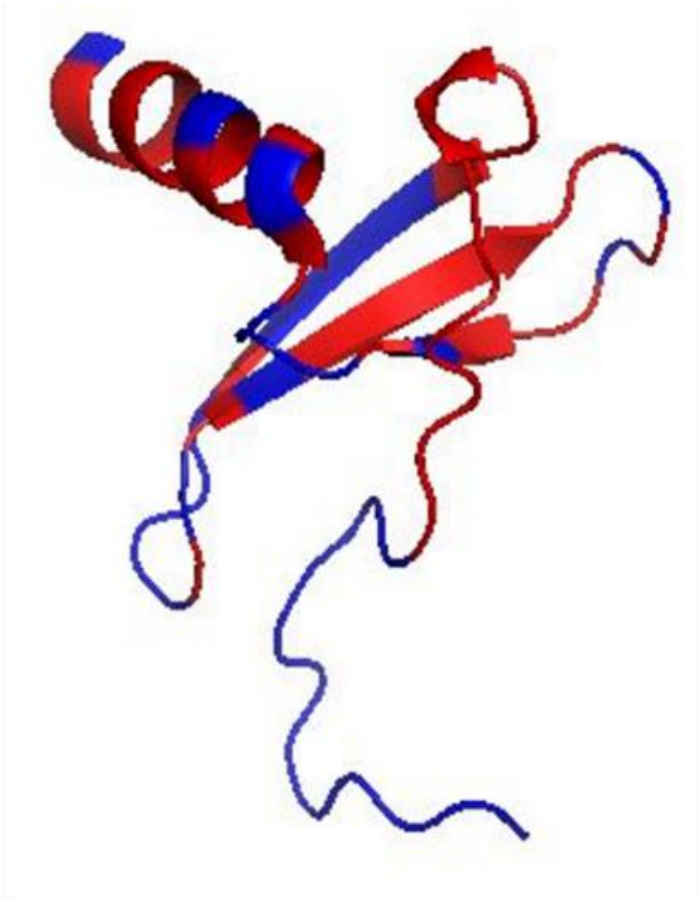

Figure 14: CXCL12 residues showing significant NMR chemical shift changes upon binding of BoxB are highlighted in red on the structure and sequence of CXLC12.

Complex between CXCL12 and BoxA

Most CXCL12 residues involved in the interaction with BoxB are also affected by the addition of BoxA to free CXCL12 (figure 15 and 16), but the

chemical shift changes were slightly different, probably reflecting interactions with residues not conserved between BoxA and BoxB that have a very similar structure but only 20% sequence identity. The CSP was generally smaller in magnitude in the BoxA complex (figure 6), possibly suggesting a weaker interaction.

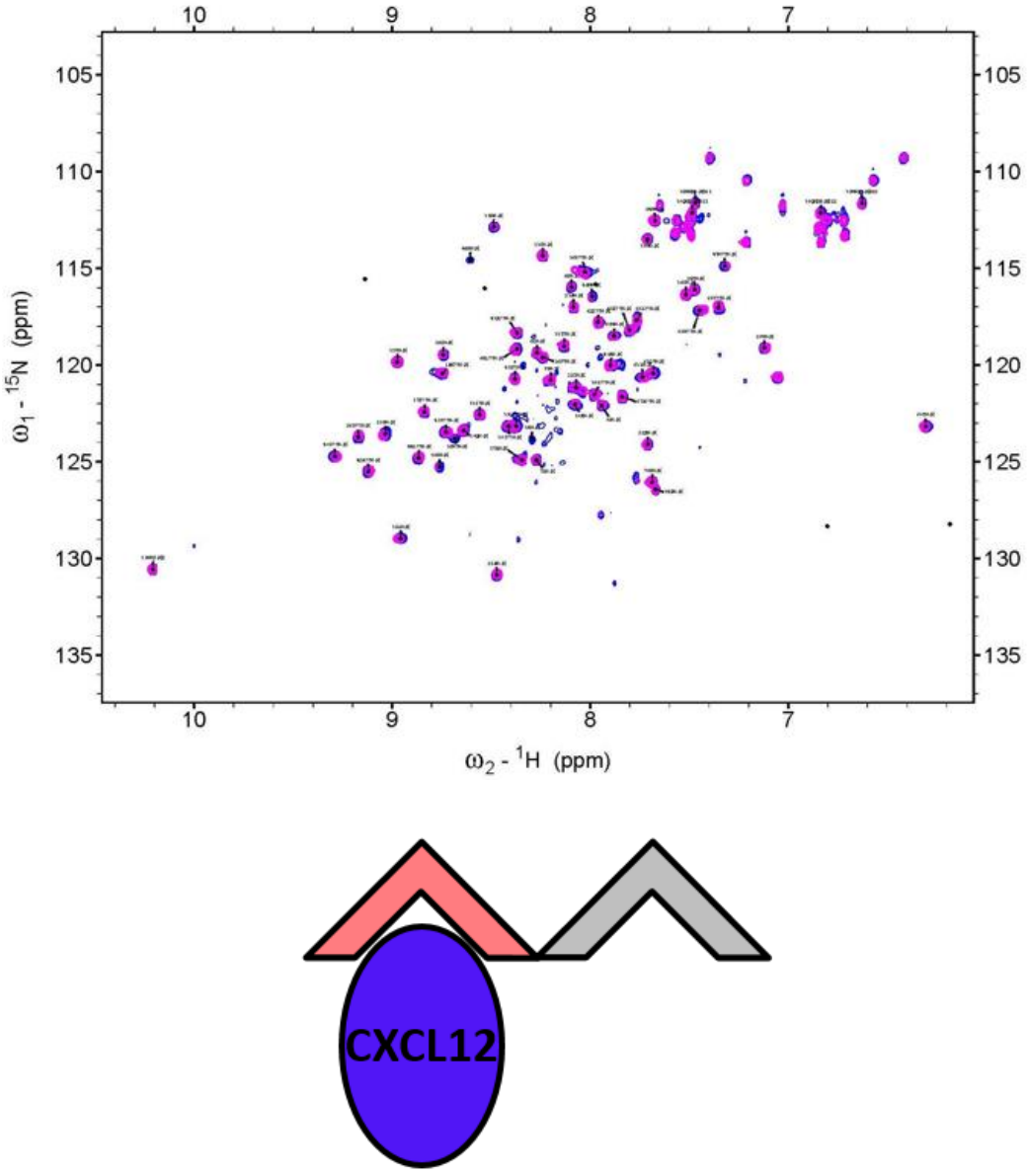

Figure 15: overlay of ^{15}N-HSQC spectra CXCL12 free in solution (blue) and in presence of 1:1 ratio of BoxA (magenta).

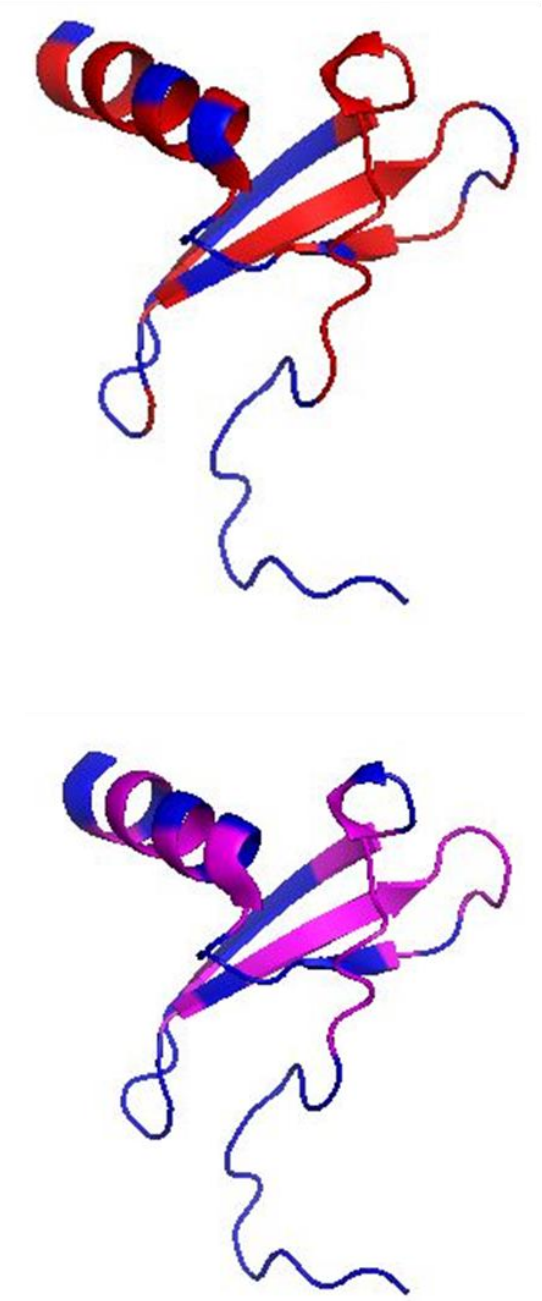

Figura 16: CXCL12 residues showing NMR chemical shift changes upon binding of BoxB (red) and BoxA (magenta) are highlighted on the CXCL12 structure (blue).

Although the interacting region of CXCL12 with BoxA and BoxB appears remarkably similar, intriguing differences are present. The signal of I51, W57 and I58 changes in the complex with BoxB but not BoxA (figure 17). These

three residues form a hydrophobic patch at the base of the C-terminal helix, which might therefore adopt a slightly different orientation in the two complexes (figure 18).

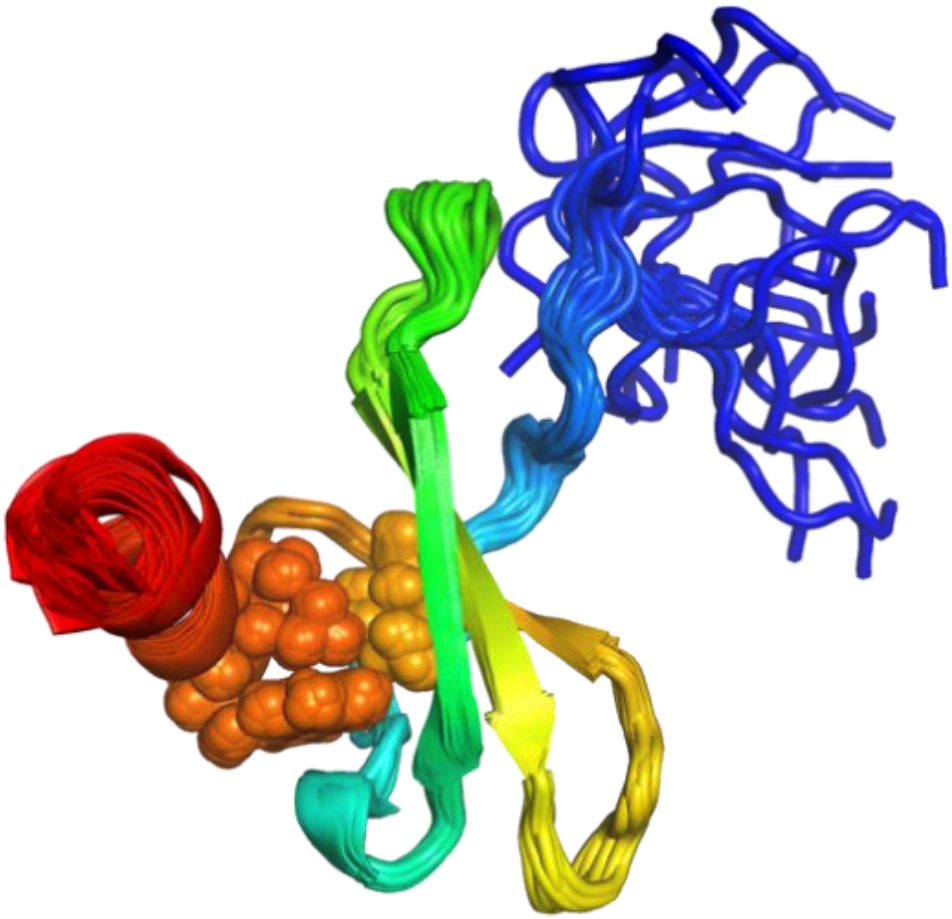

Figure 18: Three-dimensional cartoon view of CXCL12. Residues I51, I58 and W57, showing changes in the NMR signal upon addition of BoxB, but not BoxA, are shown as spheres. They form a hydrophobic patch responsible for the orientation of the C-terminal helix (red).

Mapping the interaction of CXCL12 on BoxB

In the reverse experiment we mapped the interaction of CXCL12 over BoxB. This HMGB1 domain was labelled with ^{15}N and its NMR spectra free or in complex with CXCL12 were recorded and analyzed as above.

44 out of 80 BoxB residues were affected by CXCL12 binding, confirming the interaction and pinpointing it to the concave surface formed by the BoxB helices. Intriguingly, this is the site where glycyrrhizin binds to HMGB1, which will be discussed later[126] (figure 19).

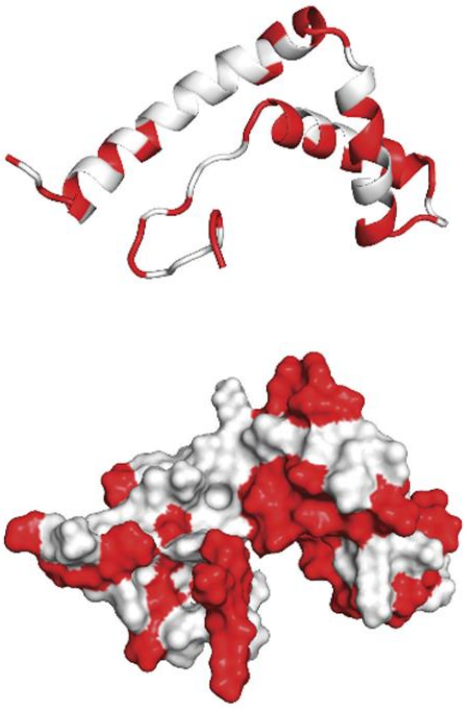

Figure 19: Cartoon and surface view of BoxB with residues showing changes in NMR signal upon addition of CXCL12 colored in red.

CXCL12 interaction with the full HMGB1 protein

The addition of full HMGB1 to CXCL12 at a 1:2 ratio causes more extensive and profound changes in the ^{15}N-HSQC NMR spectrum of CXCL12 than those induced by individual HMG-boxes. 57 of the 68 CXCL12 residues are

affected by HMGB1 addition; no signal corresponding to free CXCL12 was observed and only one set of peaks is present, i.e. each residue is present in a single conformation. This suggests a 2:1 CXCL12:HMGB1 stoichiometry with CXCL12 either assuming the same conformation in the two monomers or giving an NMR signal average of the BoxA and BoxB bound conformations.

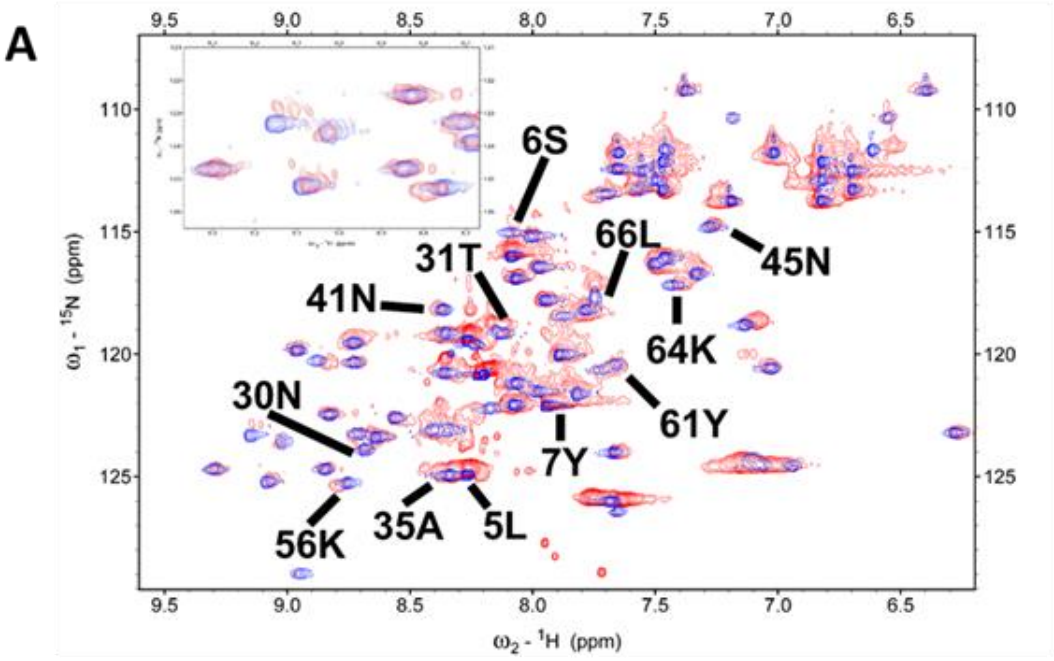

Figure 20: A) Overlay of the spectra of CXCL12 free (blue) and in complex with full HMGB1 (red) 2:1 ratio. B) Mapping of the residues showing changes

in the NMR signal upon complex formation (green) on the cartoon and surface view of CXCL12.

CXCL12 can bind to both HMG-box domains independently but has slightly different chemical shifts in the two cases. Since a single set of peaks was observed, this suggests that the heterocomplex is dynamic and the CXCL12 molecules exchange between the free state, BoxA binding, and BoxB binding (figure 20).

Figure 21 offers a comparison of the chemical shift mapping results for the binding of CXCL12 to the isolated BoxA and BoxB and to full HMGB1. All residues affected by BoxA and BoxB binding are also affected by HMGB1 binding. Additionally, residues in the N-terminus of CXCL12 and in a loop spatially close to the N-terminus are affected by binding to the full protein but not by binding to the individual boxes.

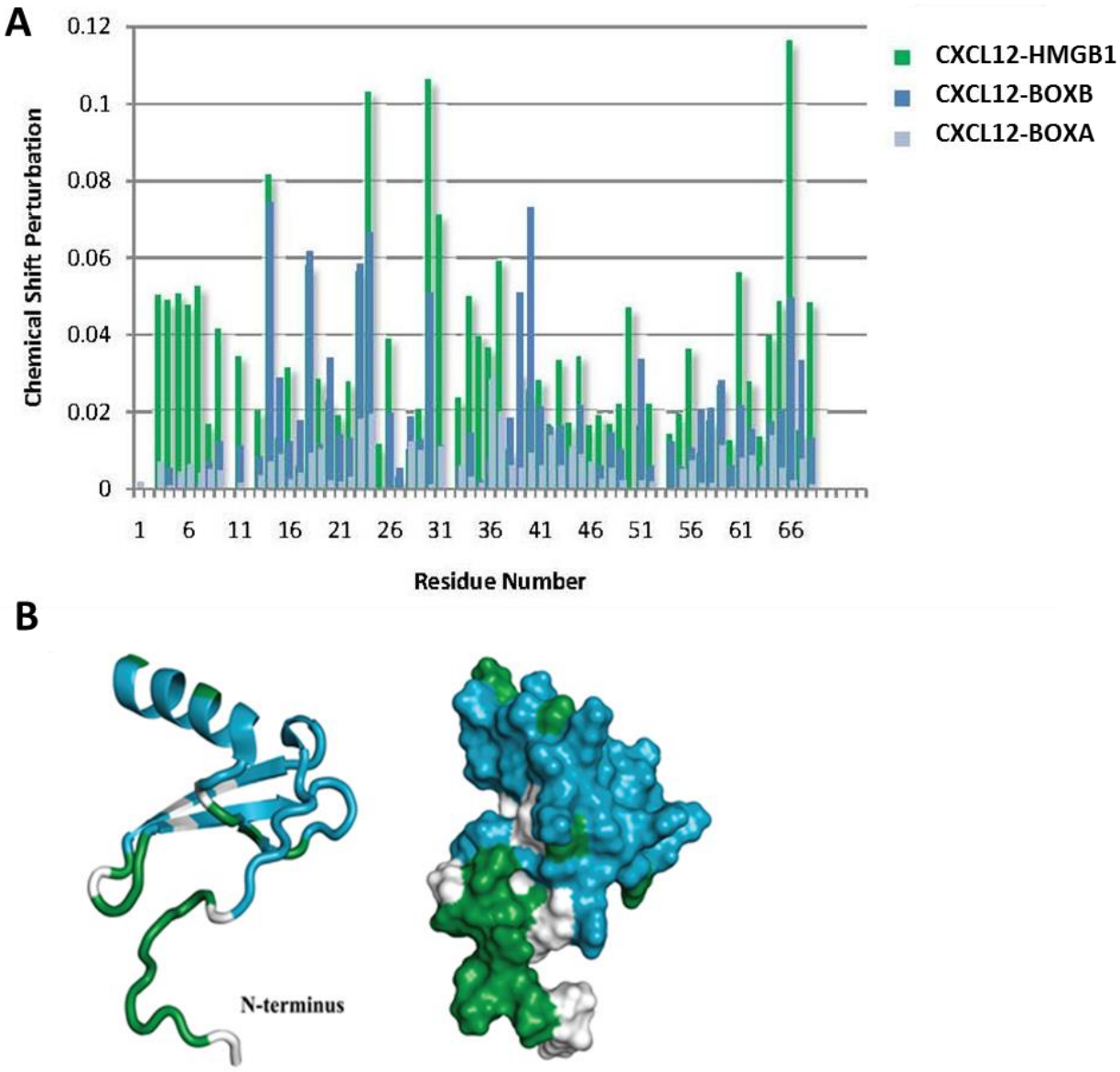

Figure 21: A) Chemical shift perturbation index plotted against the CXCL12 sequence. CXCL12 residues showing changes in NMR signal in presence of BoxB or BoxA, 1:1 ratio, are in light and dark blue. Residues affected by binding to full HMGB1 (2:1 ratio) but not to the individual boxes are shown in green. B) Mapping of the above residues on the cartoon and surface view of CXCL12. A region around the N-terminus of CXCL12 is only affected by binding to the full HMGB1 protein.

Residues 3–12 in the N terminus of CXCL12, which are directly involved in CXCR4 recognition and triggering, were affected by the binding of full-length HMGB1 but not of either BoxA or BoxB alone. The same is true for the 31–35 loop, which is spatially close to the N terminus. We know that full HMGB1 can increase the cellular activity of CXCL12 whereas the individual boxes cannot. Intriguingly, the CXCL12 residues known to interact with the cellular CXCR4 receptor (and thus triggering activity) are affected by binding to the full protein but not to the individual boxes. This suggests interesting biological consequences that will be discussed later in the thesis.

Surface Plasmon Resonance

Cross competition experiments to further probe the HMGB1/CXCL12 interface

We used Surface Plasmon Resonance (SPR) to further probe the interaction of CXCL12 with HMGB1. Detailed explanation of the SPR technology is available in the methods but, briefly, the idea is to immobilize one protein on a sensor surface and record a signal if and when a second molecule is added and interacts with the immobilized one. Analysis of the signal allows determination of association and dissociation rates as well as overall binding affinity.

Both CXCL12 and HMGB1 were immobilized on the sensor surface in a plethora of different conditions. Unfortunately, unspecific binding was always present and did not allow the reliable determination of binding properties. Unspecific binding happens if the protein flowed in solution "sticks" unspecifically to the sensor surface instead of interacting specifically with the immobilized protein. As a consequence, the recorded SPR signals do not reflect a real interaction but rather the deposition of protein on the sensor surface. It is perhaps worth reminding the unspecific signal can be reliably identified if, for instance, an SPR signal is recorded in a control channel without immobilized protein.

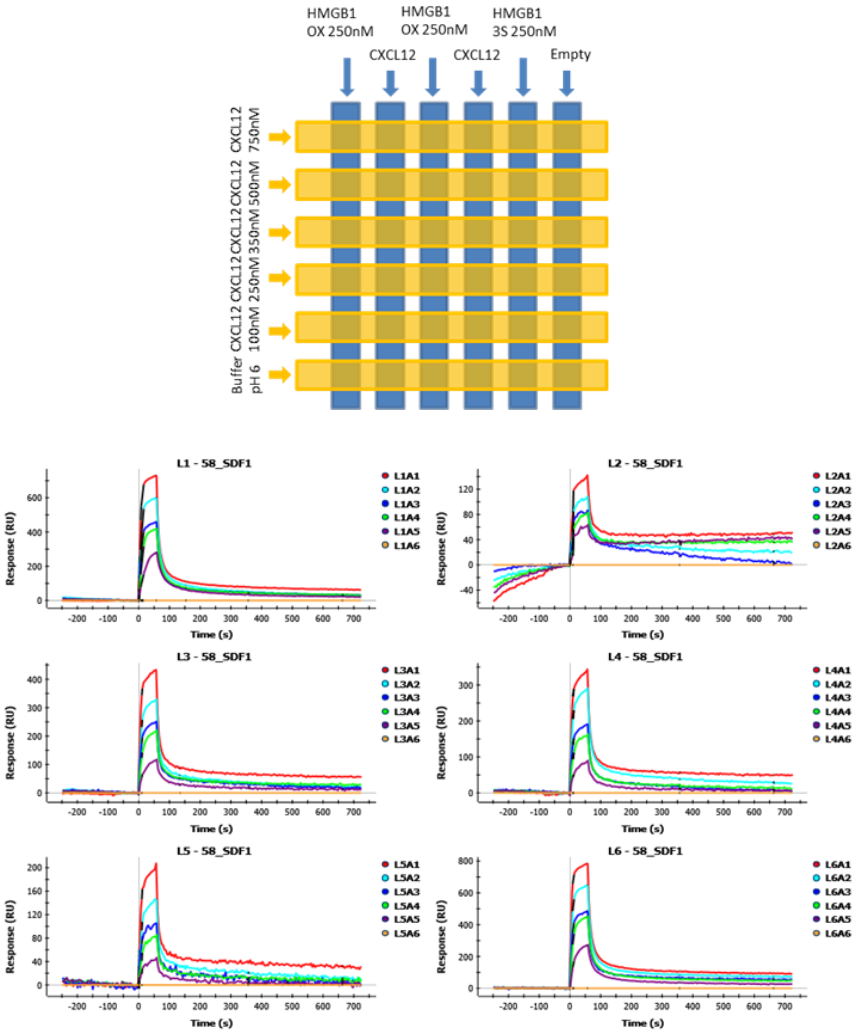

Figure 22: SPR experimental scheme (left) and sensorgrams (right) with immobilized HMGB1, either in reduced or oxidized form, and CXCL12 in solution (Phosphate buffer pH 6 in this experiment). Although a dose response is clearly present, unspecific binding of CXCL12 to the sensor surface prevents analysis of the binding properties.

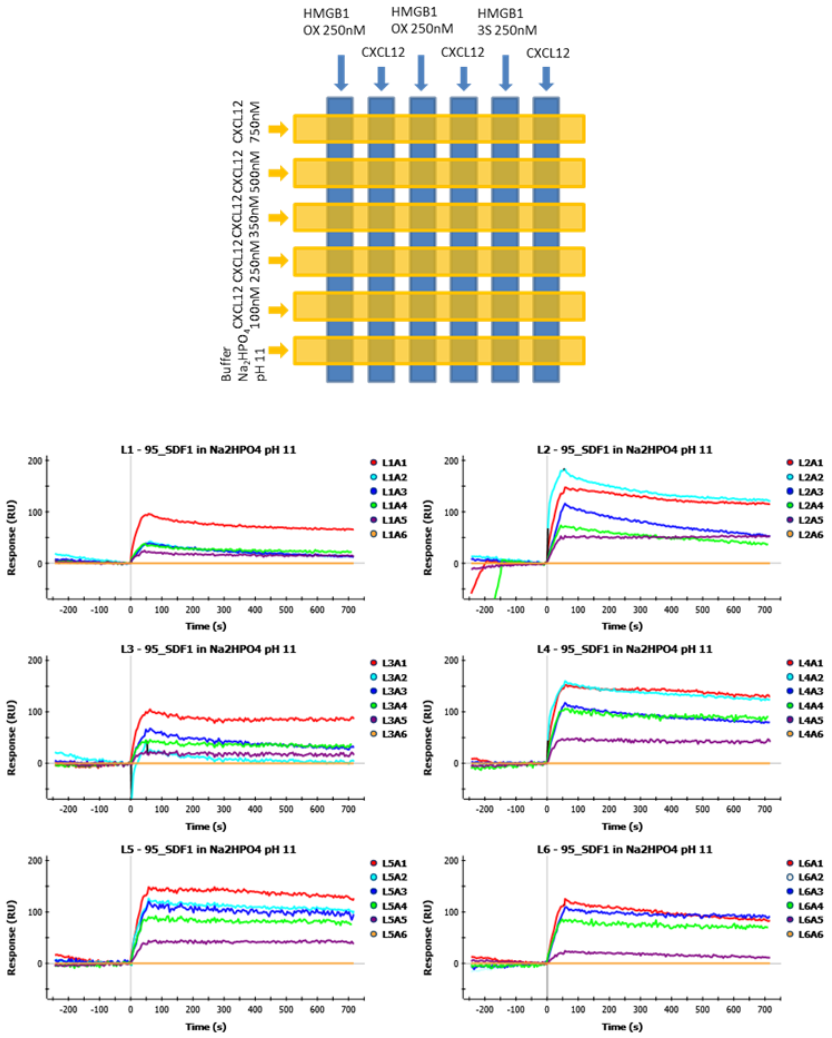

Figure 23: HMGB1 oxidized and reduced immobilized on the chip surface and CXCL12 flowed in solution at pH 11 (dimeric).

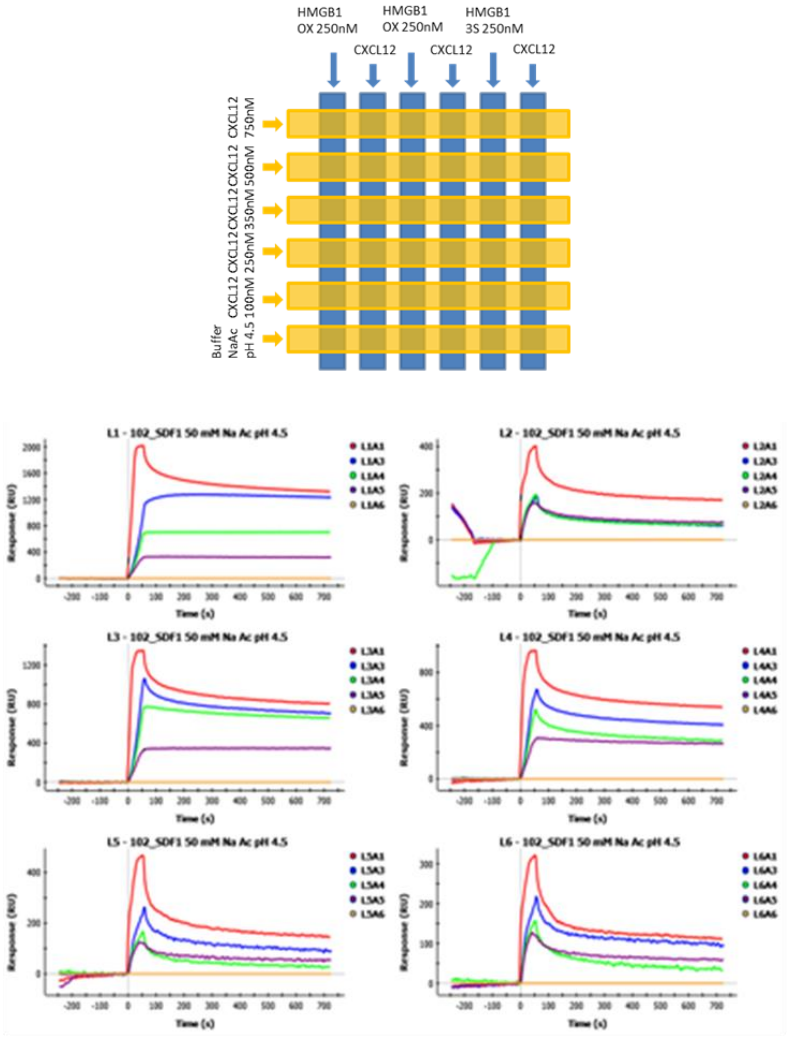

Figure 24: HMGB1 oxidized and reduced immobilized on the chip surface and CXCL12 flowed in solution at pH 4.5 (monomeric).

Since we could not directly probe and characterize the interaction with SPR, we set up a cross competition experiments to try to circumvent the unspecific binding problem. Glycyrrhizin, a component of liquorice, was shown to inhibit the chemo-attractant and mitogenic activity of HMGB1[126]. It was also shown to bind to the concave part of both BoxA and BoxB, in a region of HMGB1

that is also affected by CXCL12 binding. We hypothesized that CXCL12 and glycyrrhizin share a binding site on HMGB1 and that, therefore, can cross-compete with each other. We thus immobilized HMGB1 on a SPR sensor and then added glycyrrhizin at concentrations of 0, 500, or 5000 nM; 500 nM CXCL12 was finally added in a third and final step.

An SPR signal was detected in all cases, which is not surprising since there is unspecific binding (and thus signal) of the chemokine to the sensor surface. However, the signal intensity decreases at increasing glycyrrhizin concentration (figure 18). Since the SPR signal intensity is proportional to the amount of protein "deposited" on the sensor surface, either through interaction with the immobilized partner or by unspecific binding, a signal decrease indicates that the presence of glycyrrhizin inhibits CXCL12 binding. The SPR signal does not drop to zero because unspecific binding of CXCL12 to the sensor surface is not abolished by glycyrrhizin, whereas less and less CXCL12 can bind to HMGB1 in the presence of increasing concentration of glycyrrhizin, reflected by the decreased SPR signal. The experiment was obviously repeated at least 3 times in each condition tested and several reference channels were used, for instance flowing CXCL12 or glycyrrhizin over a sensor surface in the absence of HMGB1 (figures 22,23,24,25).

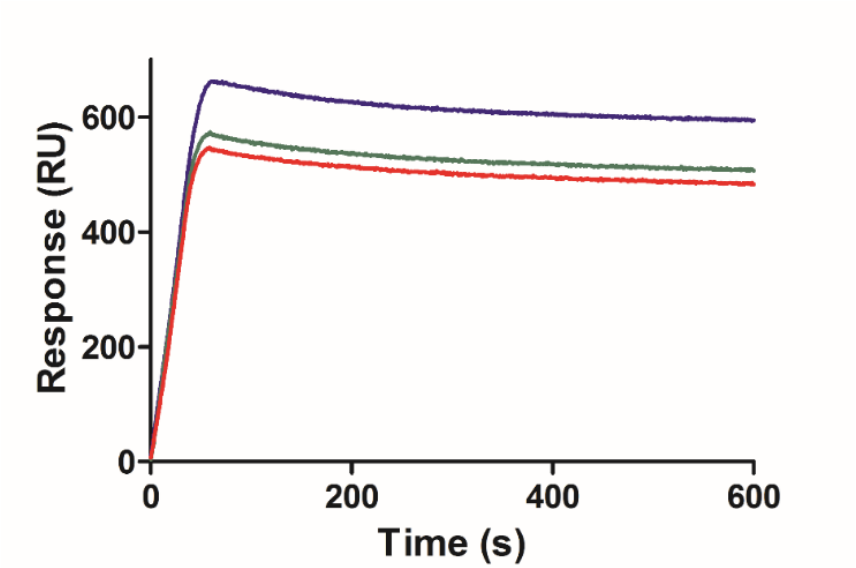

Figure 25: SPR sensorgrams of the interaction between HMGB1 and CXCL12 in the presence of increasing concentration of glycyrrhizin. HMGB1 was immobilized on the SPR sensor surface; glycyrrhizin was added at 0 (blue), 500 (green) or 5000nM (red). Finally, 500nM CXCL12 was added to each channel, recording the SPR response shown in the image. Glycyrrhizin inhibits CXCL12 binding, reflected by a decrease in SPR signal. Complete loss of signal is not achieved because unspecific binding of CXCL12 to the sensor surface cannot be eliminated.

Computational Docking

We used computational docking, the process of predicting the structure of a complex starting from the separate structure of its individual components[47], to obtain a three-dimensional atomic model of the CXCL12/HMGB1 complex. Experimental structures of both proteins are publicly available (PDB ID. 2KEE and 2YRQ).

Protein-Protein docking starts with the two molecules separated by typically 25Å and then brought together by the preferred algorithm. Translational and rotational movements generate "poses" whose energetic score is evaluated by apt scoring functions. The process stops after a pre-determined number of "moves" or when the scoring function does not improve after subsequent moves, indicating that the system has reached an energy minimum.

Each docking run generates thousands of models (decoys) of the complex and evaluates them in term of an empirically determined scoring function that attempts to represent energy states (pseudo-energy). Simply put, each hydrophobic and electrostatic interaction, salt bridge, hydrogen bond etc. contribute to a global score; statistical and empirical considerations such as the degree of conserved residues in the interface can also be taken into account. The assumption is that the structure having the lowest energy is also the biological one, although this is not always true.

When searching for the correct binding orientation, the two molecules are allowed to move and the score is assessed after each step. A Monte-Carlo minimization protocol retains some conformation with higher energy in an attempt to overcome local energy minima that don't correspond to the global minimum. The movement is stopped after a predefined number of steps or

when the score does not improve further. The conformation parameters changes in different docking stages and algorithms, which may be divided in three general classes: (i) only the relative position of the docking partner is changed; (ii) the relative position and the sidechain conformations are changed; (iii) the backbone conformation is altered in addition to the above.

The simplest case is called rigid body docking because the conformation of the starting structures is not altered at all during the docking process and the scoring function only needs to account for the intermolecular interactions[72,73]. Rosetta-Dock, the algorithm used in this thesis, has a first rigid body phase in which sidechains are removed, but in a second phase they are re-introduced and their orientation is optimized[74,127].

The simulation of backbone movements that often happen upon formation of biological complexes, instead, remains a daunting task for docking, which has a very high failure rate when molecules undergo significant conformational changes upon binding.

A plot of the score of each decoy (scoring plot) is often used to obtain a visual representation of a docking run (figure 26). The presence of a "scoring funnel" indicates that the process is finding several similar structures with low energy, which is usually considered an indication of accuracy (care must always be taken, however).

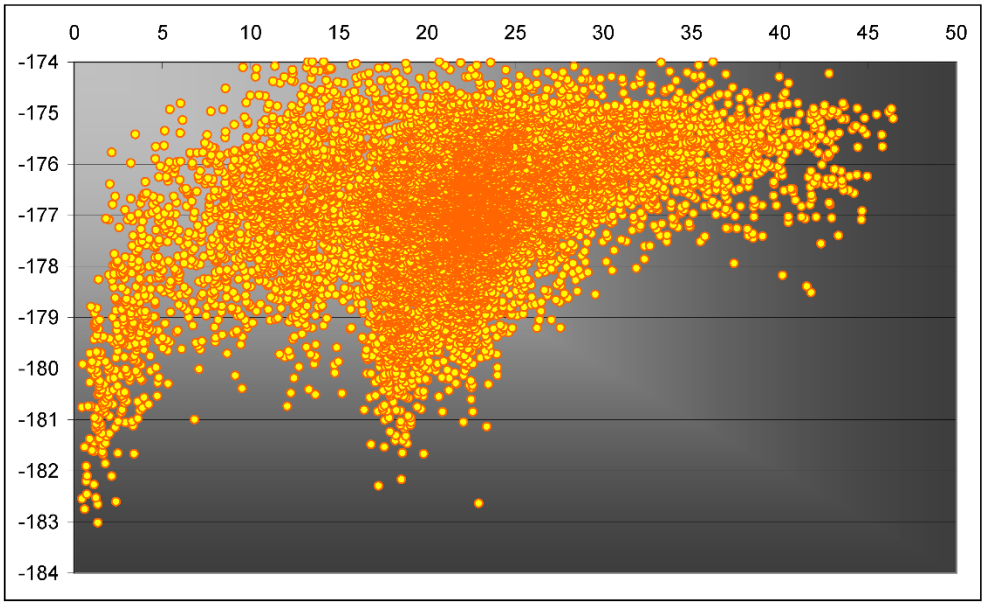

Figure 26: Scoring plot. Each dot corresponds to a docking model obtained during the simulation. Lower values for the scoring function (pseudo-energy, y-axis) indicate solutions considered more accurate by the docking algorithm. The x-axis show structural difference (RMSD) to a reference structure; similar RMSD values tend to indicate similar structures.

Although docking algorithms are constantly improving, scoring function often fail to discriminate accurate from inaccurate solutions. Experimental data from NMR studies, like the mapping we conducted on the HMGB1/CXCL12 complex, can provide a huge help to discriminate accurate solutions[128]. Indeed NMR chemical shift perturbation (CSP) analysis identifies the region of the proteins affected by binding during complex formation and thus likely to be at the interface, although allosteric effects might cause CSP changes as well.

12

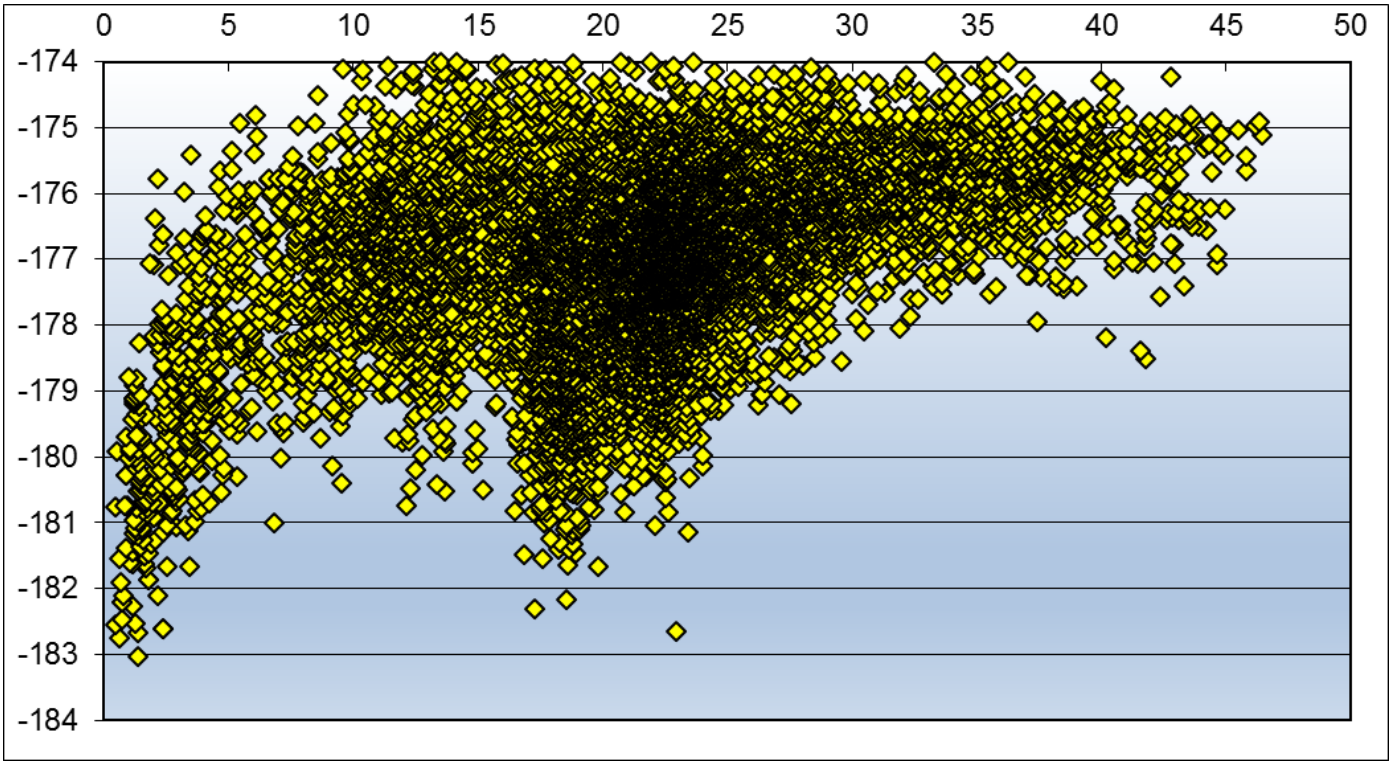

Figure 27: Scoring plot of the docking between CXCL12 and BoxA.

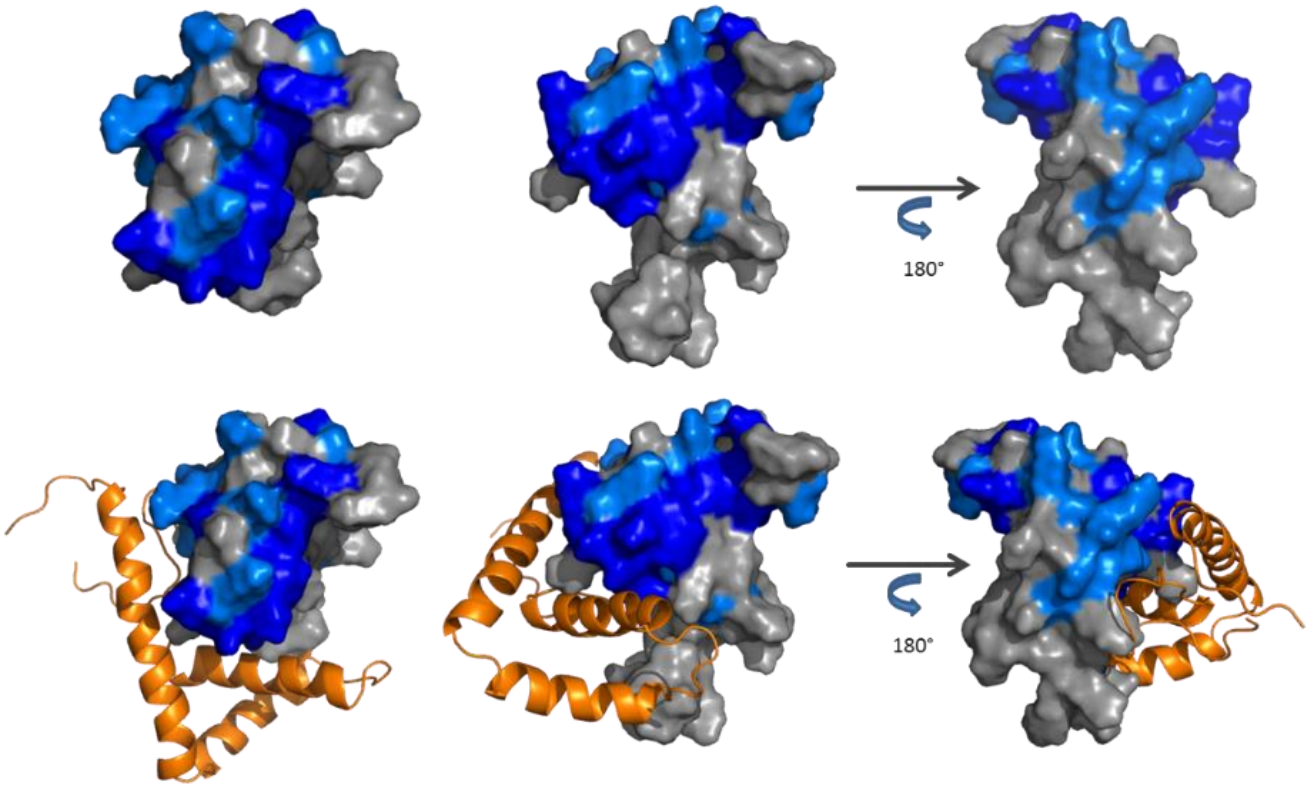

Figure 28: The best model obtained for the complex between CXCL12 and BoxA. Top: surface view of CXCL12 with the binding region, according to NMR chemical shift mapping experiments, shown in blue. Bottom: docking model of the complex between BoxA (orange) and CXCL12. There is no good agreement between computational and experimental data.

The docking simulation of CXCL12 in complex with BoxB was performed with different approaches because BoxB (in the construct used for the NMR

experiments) contains an unstructured linker that connects it to BoxA. First of all the docking simulation was run between CXCL12 and "full" BoxB, but the unstructured linker interfered with the position of CXCL12. This situation is probably unrealistic, since the linker would be free to move out of the way in reality but is not allowed to do so during the simulation. For this reason, a second set of docking simulations was performed removing the tail from the docked construct (figure 29).

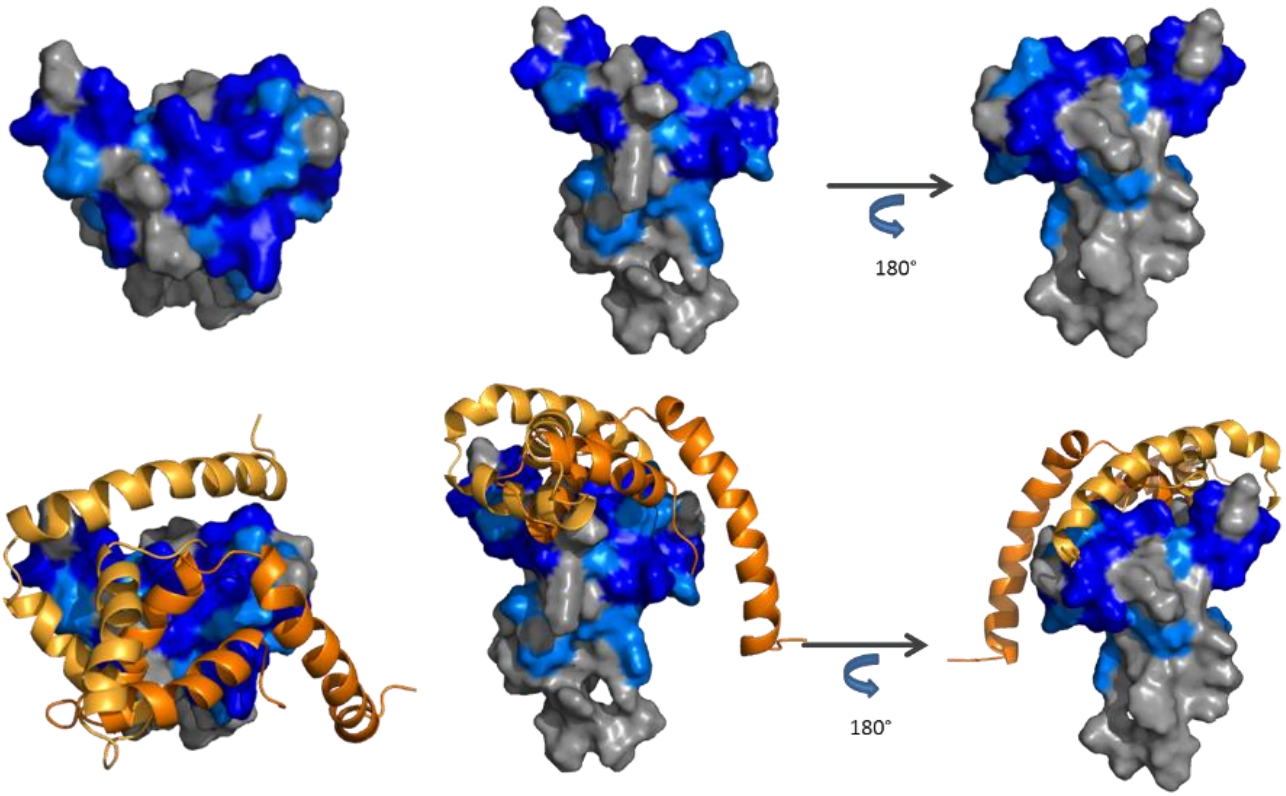

Figure 29: The best model obtained during the simulation between CXCL12 and BoxB. Top: surface view of CXCL12 with the binding region obtained with NMR chemical shift mapping experiments in blue. Bottom: overlays of

two docking models with BoxB shown in orange. In both models there is good agreement between docking and experimental data.

Unfortunately, the binding region experimentally defined on the CXCL12 structure is vast and covers most of CXCL12; as a consequence it was not possible to define a unique solution; on the contrary, several docking models equally satisfied the experimental data and are, therefore, considered equally accurate (figure 30).

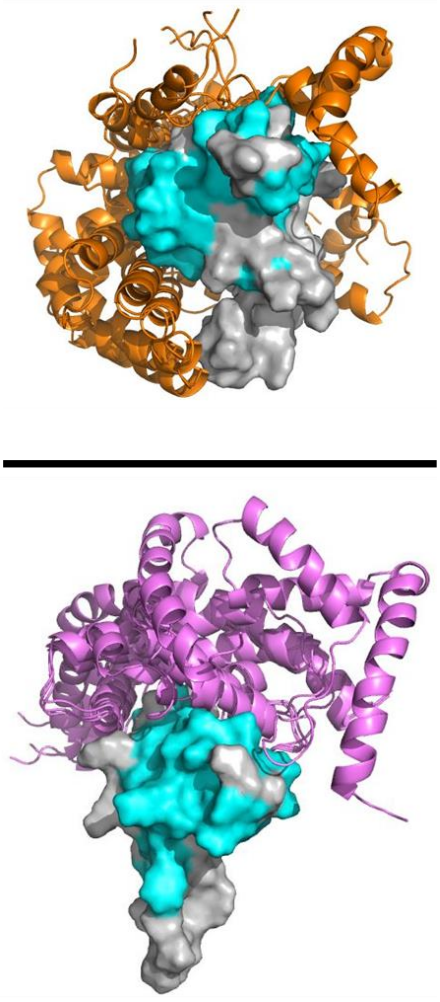

Figure 30: Top: Docking between CXCL12 (grey) and BoxA (orange), with the NMR defined binding region in cyan. Several docking models are overlaid. Bottom: as above for the CXCL12/BoxB complex.

Computational docking was unable to define a reliable atomic model for the complex between HMGB1 and CXCL12. One possibility is that HMGB1 and/or CXCL12 might undergo conformational changes upon binding, something that computational docking cannot effectively simulate. At the same time, the docking models with best computational score and the agreement with the NMR mapping data suggest that both BoxA and BoxB bind to the globular region of CXCL12 using their DNA-binding kink.

Discussion and conclusion

Cellular experiments indicate that the cellular activity of CXCL12 is enhanced by the presence of HMGB1 protein. A direct interaction between the two entities could, however, not be established.

Our NMR studies prove that both BoxB and BoxA use the kink in their structure, corresponding to the DNA binding site, to bind CXCL12. The experiments define the CXCL12 region responsible for HMGB1 binding, which involves the globular domain and is identical in the complexes with the two Boxes. In our hypothesis two molecules of CXCL12 interact with one molecule of HMGB1 by direct binding to the individual HMGB1 subunits. Binding involves the globular part of CXCL12, composed of one alpha helix and 3 beta strands, and the concave region of the HMGB1 subunits, which is also the glycyrrhizin binding site.

CXCL12 is known to dimerize in certain conditions and an experimental structure of a CXCL12 dimer is available, although this is somehow artificial since the two subunits were covalently cross-linked[129]. Our data excludes that such dimer is formed in the presence of full HMGB1. NMR mapping shows that the dimer interface residues are not affected by HMGB1 binding. This would not be the case if that dimer were formed upon binding (figure 31).

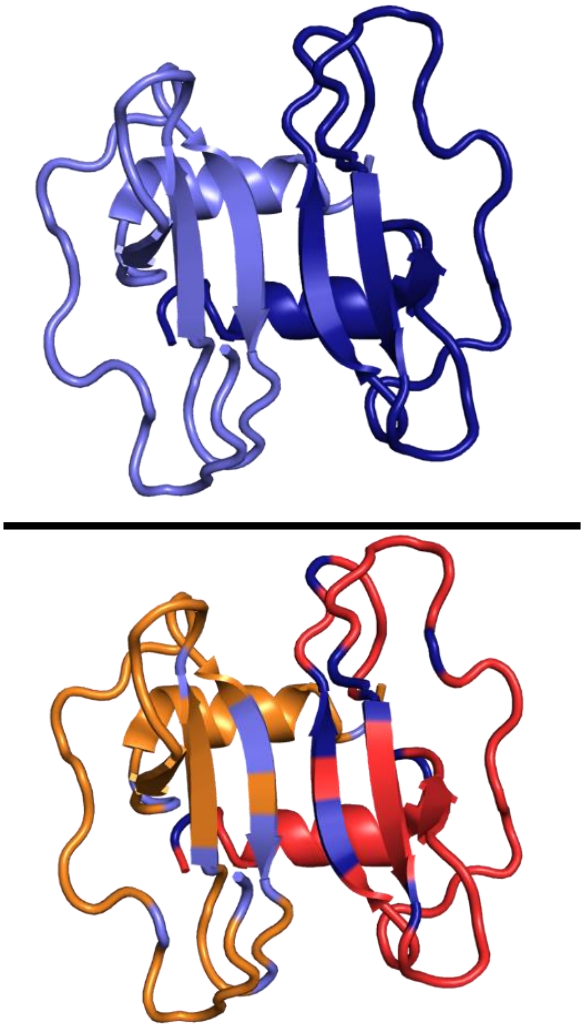

Figure 31: Dimeric conformation of CXCL12 (PDB ID 2K01), the two monomers are highlight in dark and light blue on the top while on the bottom orange and red indicate the residues affected by the formation of the CXCL12/HMGB1 complex.

What we propose, instead, is a model with the N-terminal tail of the two CXCL12 monomers arranged in a parallel fashion, as in figure 26. In such configuration the CXCL12 N-terminus does not contact BoxA or BoxB, in agreement with NMR mapping showing that they are not affected by binding. Upon binding to full HMGB1, instead, the two N-terminus regions might

interact with each other, in agreement with the NMR mapping data showing that they are indeed affected by full HMGB1 binding (figure 32).

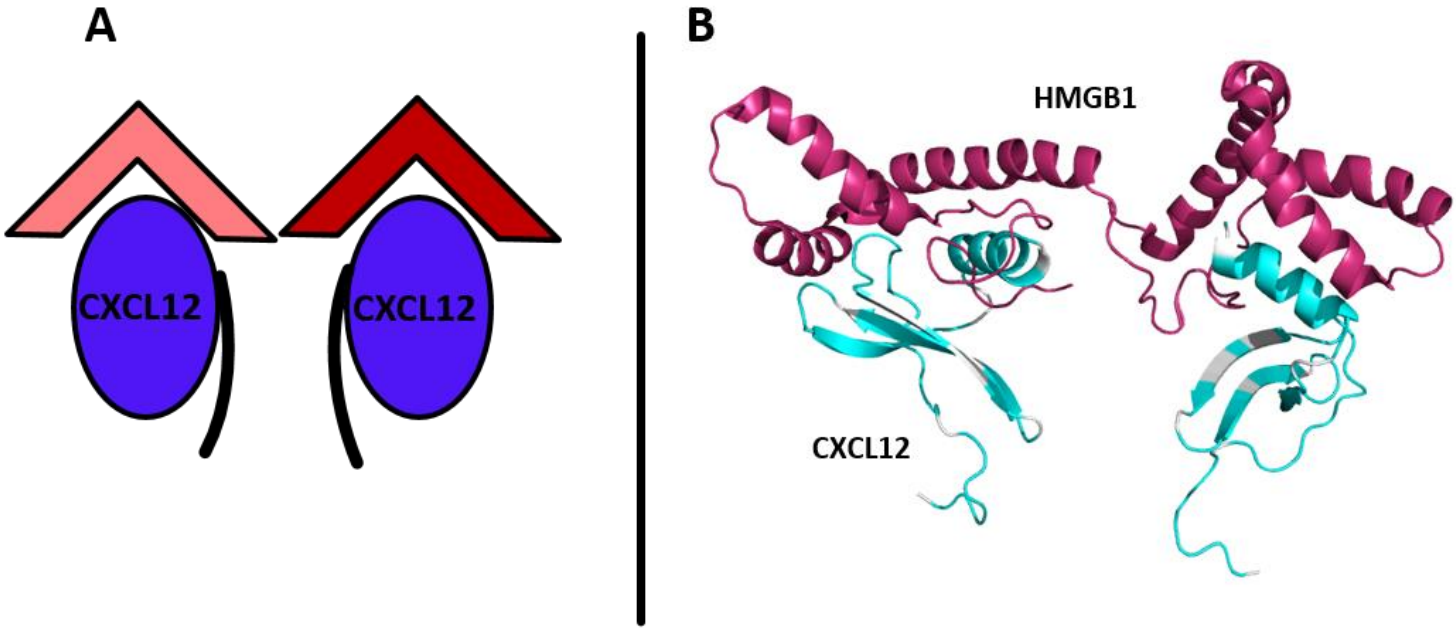

Figure 32: Hypothesis of interaction between CXCL12 and HMGB1. Two molecules of CXCL12 bind one molecule of HMGB1, the globular head of CXCL12 binds to the kinks of the HMGB1 sub-units.

It is important to note that the presence of either Box is sufficient for binding to CXCL12 but not to enhance its cellular activity which is, instead, increased in the presence of full HMGB1. NMR shows that the flexible N-terminus residues, as well as a spatially close loop, are affected only by binding to full HMGB1. Residues 3-12 of CXCL12 residues are required for CXCL12 to interact with its cellular receptor, CXCR4, and to trigger its cellular activities. The modality of interaction of CXCL12 with its receptor is known and it is well explained by the "Fish and Rod" model[130]. Initially residues in the β-

sheet/α-helix of CXCL12 interact with the extracellular loops of the CXCR4 receptor. This facilitates the anchoring of CXCL12 to CXCR4 with the CXCL12 N-terminus remaining highly mobile (as shown by NMR dynamics studies[130]). This mobility allows the N-terminal residues to "probe" the binding cavities buried within the transmembrane helices of CXCR4. In a second step of interaction the CXCL12 N-terminus is inserted in the transmembrane helices, presumably immobilized, triggering conformational changes in CXCR4 that induce G-protein signaling. Blocking the flexible N-terminus in a single, probably rigid, conformation requires a certain entropic cost. In our proposed model there are three partners: CXCL12, HMGB1 and CXCR4. Binding to full HMGB1 puts the CXCL12 N-terminus in a conformation ready to interact with CXCR4. The energy spent to lock into such orientation is paid by HMGB1 binding. As a consequence, binding of the HMGB1/CXCL12 complex to CXCR4 would be energetically more convenient than binding of CXCL12 alone, explaining the increased cellular activity of this chemokine in the presence of HMGB1. There are strong suggestions that CXCR4 dimerizes on the cell membrane; the presence of two CXCL12 molecules bound to HMGB1 could also favor interaction with the receptor dimer, either by putting CXCL12 in the right conformation for interaction (as above), or simply by rendering the binding of the second unit more favorable after the first one has locked onto CXCR4, effectively increasing the local CXCL12 concentration (figure 33).

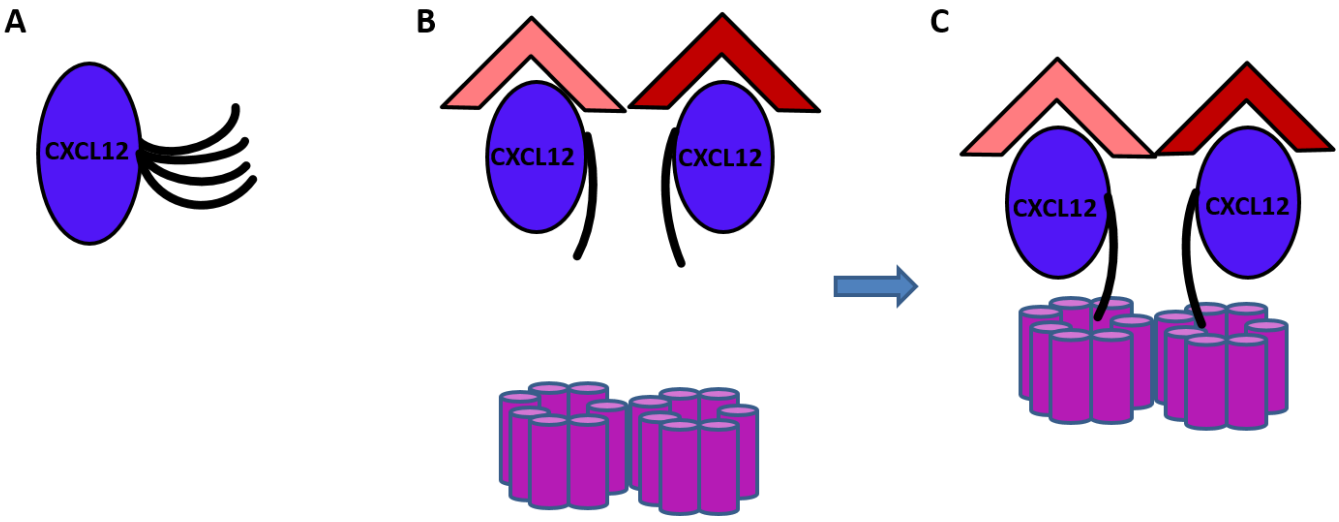

Figure 33: CXCL12 free in solution (A) with the highly mobile N-terminus shown as lines in multiple conformations. In our hypothesis HMGB1 binds two molecules of CXCL12 and locks the N-terminus in a single conformation (B) ready to interact with CXCR4 (C).

Section IV

Rational engineering of human antibodies through experimentally validated computational docking

Abstract

Dengue virus (DENV) is a mosquito-borne flavivirus that in recent decades has become a major international public health concern[146].

The virus is found in tropical and sub-tropical regions, predominantly in urban areas, and an estimated 3.5 billion people, or half the world's population, are at risk of infection, with the number constantly increasing[146,149].

The long term aim of my work is to understand why some antibodies are able to neutralize only some Dengue serotypes but not others and why most antibodies facilitate subsequent infection by other serotypes (Antibody Dependent Enhancement) , leading to the insurgence of Dengue Hemorrhagic Fever (DHF)[149].

In order to reach this goal, the main objective of this project is the structural and biophysical characterization of new neutralizing and/or enhancing antibodies from Dengue virus recovered subjects.

In particular, we want to compare the binding of the same antibody to the four Dengue serotypes and the binding of different antibodies to the same serotype, therefore it is necessary to determine several different antibodies-DIII structures.

Traditional experimental techniques like X-ray crystallography or NMR spectroscopy can give us very precise and accurate results but are long and expensive methods. If we want to characterize a large number of Ab-Ag complexes, instead, we cannot rely on these methods but need a faster

approach. Computational docking might be such an approach and even if its results are not going to be as accurate and precise as an x-ray structure, sacrificing precision for speed might be worthwhile if looking for general trends in a large panel of structures.

In a typical docking simulation the structures of the individual partners are brought together by the chosen algorithm. While we know the experimental structure of DIII, we need to model the structure of the antibodies of interest. The presence of a conserved framework region, and the fact that most CDR loops have access to a limit number of canonical structural classes make antibodies particularly suitable for homology modeling prediction. 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.

In order to overcome this problem, in our approach we obtain 11 different models of the antibodies and we dock each of them independently with Dengue, in order to explore the effect of different loop conformations and to limit the chances of starting with a wrong antibody model.

Introduction

Dengue virus (DENV), a mosquito-borne flavivirus like West Nile Virus and Yellow fever, is the causative agent of dengue fever, one of the most significant emerging disease challenges to global public health[131,132].

Dengue virus is responsible for ~100 million annual human cases, including 500,000 hospitalizations and 20,000 deaths with an economic burden rivaling that of malaria. Although DENV has been mainly restricted to the tropical region, both its epidemic activity and its geographic expansion are increasing as travel, urbanization, and climate changes create favorable conditions for vector and virus dissemination, with an estimated 2.5 billion people at risk of infection[133-135].

At present, neither a vaccine nor anti-viral drugs are available, prompting research initiatives aimed at understanding the molecular and cellular virology, genomics and evolution of this important virus. The effort to find a vaccine has been hampered by the presence of four different dengue serotypes (DENV1–4) and by a poorly understood process, antibody-dependent enhancement (ADE). Antibodies raised against a previous Dengue infection facilitate subsequent infection by a different serotype and lead to

Dengue Hemorrhagic Fever (DHF), a more severe form of the disease with fatality rates that can surpass 20% but fall below 1% with appropriate supportive care[136-138]. This feature complicates the task of finding a vaccine,

since a vaccine that would not protect equally against all four serotypes would actually contribute to the emergence of Dengue Hemorrhagic Fever[139].

In order to understand a good approach is to Characterizing and comparing the binding of several antibodies to the viral surface protein of the four different Dengue serotypes is expected to improve our understanding of the molecular basis of neutralization and ADE. In particular, this strategy allows us to answer two main questions regarding the immunological properties of Dengue virus:

- Do all the antibodies that confer protection from disease bind to the same region of Dengue? If this is the case than we could identify such binding region and ideally exploit this knowledge to design a vaccine mimicking it.

Are there common properties amongst the antibodies capable of neutralizing a given serotype? If so we may introduce such properties in new antibodies lacking them, designing new molecules or optimizing the existing ones for passive immunization. The Dengue virus particle is a sphere with a diameter of approximately 500 Å and is composed of a single, positive-strand RNA genome packaged by a virus capsid protein in a host-derived lipid bilayers[140,141].

The viral surface is formed by 180 copies of an envelope glycoprotein called E protein. E protein forms a dimer and is the dominant target of the response

consisting of antibodies against Dengue. Each monomer is subdivided in three domains: DI, DII and DIII. The central structural domain (domain I) contains the N-terminus and is flanked on one side by an elongated dimerization domain (domain II) which contains the fusion peptide at its distal end. On the other side of domain I is domain III, an immunoglobulin (Ig)-like domain thought to contain the receptor-binding sites[141].(Fig. 34) The latter is the favorite target of neutralizing antibodies against Dengue and is also the most variable region amongst different Dengue serotypes, making it an attractive target for our structural immunology studies[142-146].

Since the cellular effect of an antibody (Ab) starts, in ultimate analysis, with specific atomic interactions between the Ab itself and the pathogen, a structural comparison of several different Ab/DIII complexes could provide insightful clues on the features needed for effective DENV neutralization and enhancement[147]. Ideally we want to compare the binding of the same Ab to DIII of different serotypes and the binding of different Abs to the same serotype. If we want to observe many structures, however, we need an approach faster than traditional experimental techniques such as X-ray crystallography or NMR spectroscopy. Computational structural biology has the potential to be such approach.

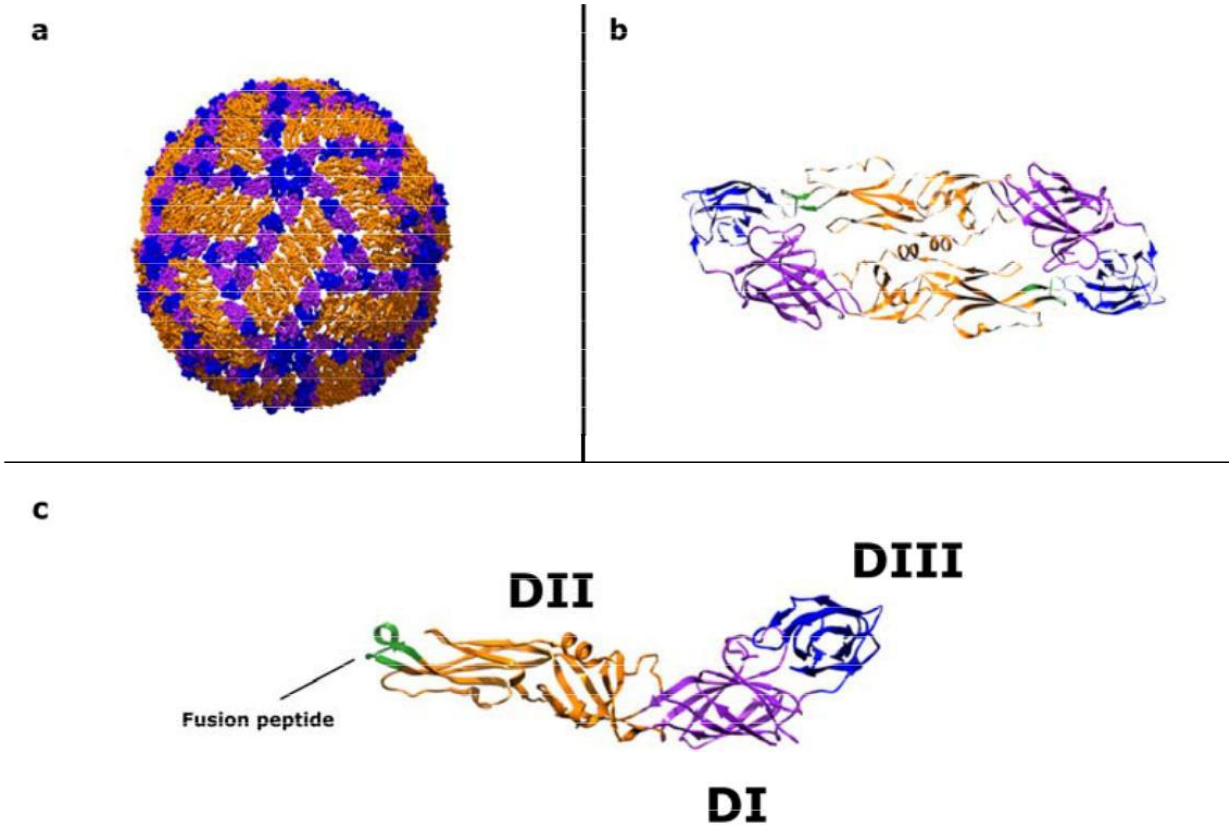

Fig.34 Three-dimensional structure of Dengue virus. (a) the entire surface of the virion formed by 180 copies of envelope protein (E protein) that forms dimers, shown in panel (b). (c) The E-protein monomer is formed by three domains called DI, DII and DIII. DII contain the fusion peptide, in green.

Results

The epitope of antibody DV32.6 on Dengue1 DIII

We characterized the binding of DV32.6, the same antibody described in the JMB paper[35], to DENV1 and here compare it to DENV4.

Thirty-two residues of the E protein domain III of Dengue serotype 1 are affected by the binding of DV32.6. The epitope is almost identical to that of DENV4-DIII, at least at the residue level. Aligning the structure of DENV1/DIII with the one of DENV4/DIII and mapping the residues that experience a shift in their NMR signal on the surface of DENV1/DIII, as we done for DENV4/DIII, we can observe a slight difference position of the epitope region between DENV1 and DENV4, as shown in figure 35.

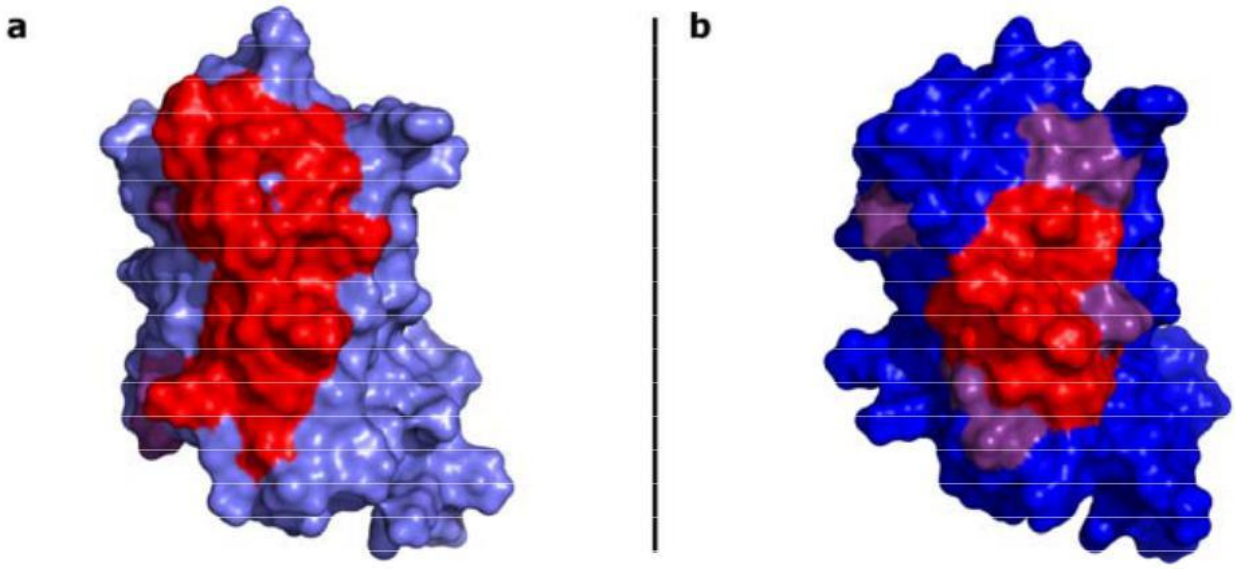

Fig. 35 Residues whose NMR signal changes upon interaction with DV32.6 are mapped on the surface of DENV1/DIII (a) and DENV4/DIII (b). DIIIs are colored in different shades. Residues showing significant chemical shift changes are in red, small changes in purple.

Binding of DV32.6 to DENV1 E protein domain III

Using RosettaDock software we perform docking simulations using the experimentally obtained structure of DENV1/DIII and each of the eleven structures of DV32.6 antibody obtained with RosettaAntibody and PIGS servers. As for the case of DENV4/DIII, not all of them give a docking result that is in accordance with NMR experimental data. The models of the complex obtained using Models 4 and 8 show no violation according to the NMR and SASA criteria and their ensemble is accepted as final result (Fig. 36).

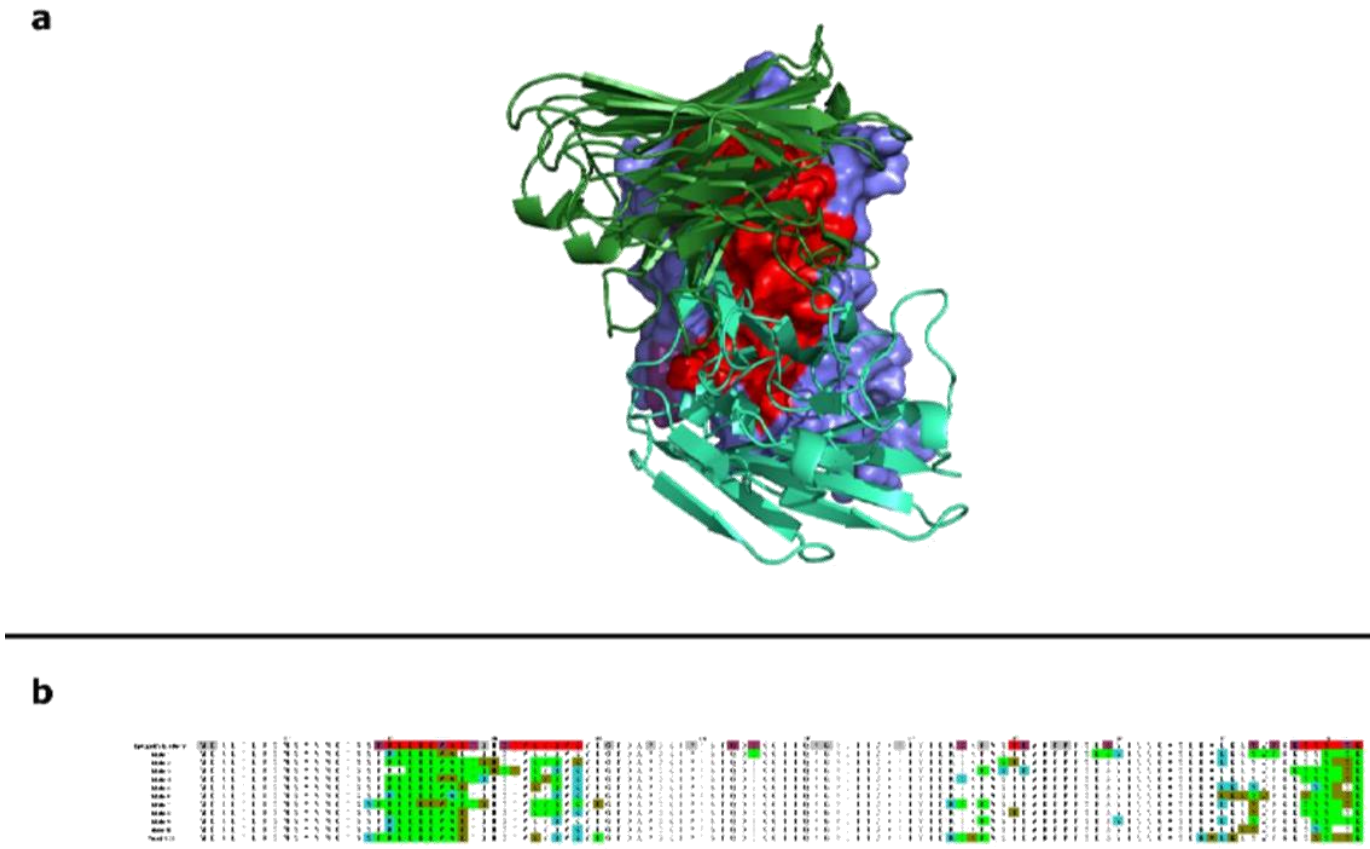

Fig. 36 Docking multiple models of the Ab. (a) Superposition of all DV32.6 models showing no violation; DENV1/DIII and the epitope are colored as in the previous figure; the Abs are shown as cartoon. The heavy and light chains

are in dark and light green, respectively. (b) Top line: the NMR epitope is mapped on the sequence of DIII. Residues with significant chemical shift changes are in red, while residues with small changes are in purple. Residues for which no NMR mapping information is available, either because of overlap or because of lack of assignment, are in gray. Bottom lines: interface residues are highlighted for each model. Residues at the interface according to both the NMR and SASA criteria (see attached paper) are in green. If they are at the interface according to the NMR but not the SASA criterion, they are shown in dark green. If they are at the interface according to the SASA but not NMR criterion, they are in cyan. Models 4 and 8 have no violations.

In contrast to DENV4, where DV32.6 utilizes the light chain CDR loops plus the H3, the antibody contacts DENV1 using also the H1 and H2 loops (D33 in H1 loop has a ionic interaction with K400 while S52, S55 and Y57 in H2 loop with S397, S398 and K400 respectively) (Fig. 37).

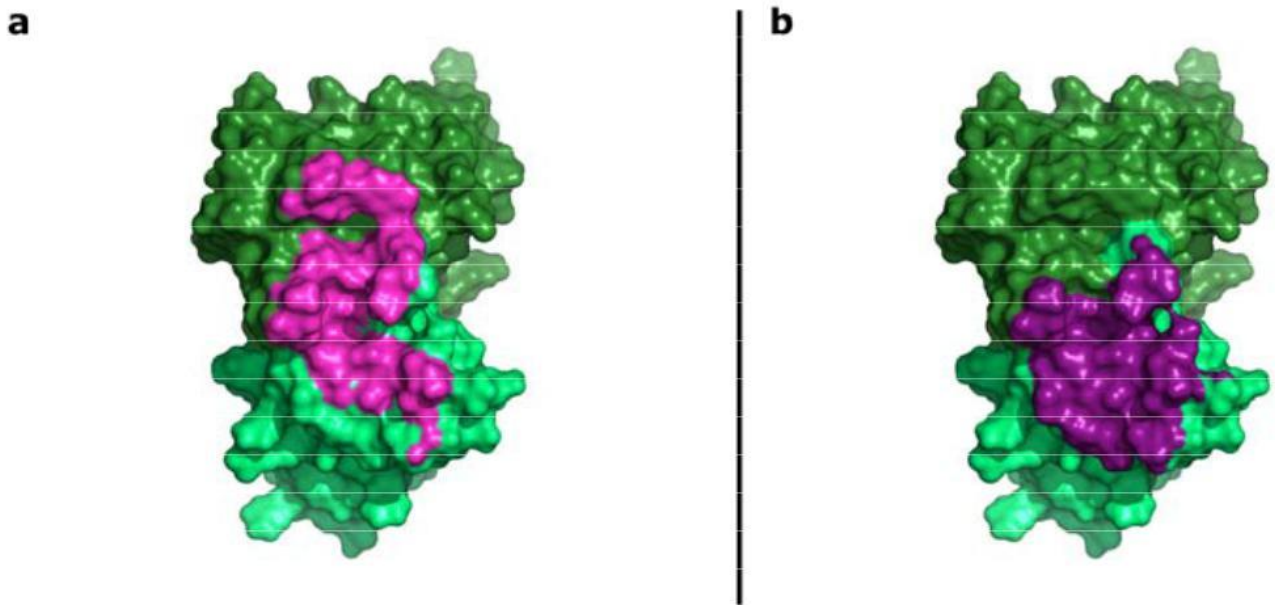

Fig. 37 Footprints of DENV1/DIII and DENV4/DIII on DV32.6. The surface of DV32.6 is shown, with heavy and light chains colored in dark and light green, respectively. Residues that make contacts with the DENV1/DIII are colored light-violet (a), while the ones that interact with DENV4/DIII are colored in purple (b).

The center of the binding is characterized by interactions between D50 and D51 of the light chain of DV32.6 with three lysines in the center of the epitope. In particular, the side chain oxigen of D50 makes a hydrogen bond with the NH_3 group of K310 while the two oxigens of D51 bridge the NH_3 groups of two lysines on the antigen, K307 and K325. Furthermore, the side chain oxygen of D102 on the Ab makes an hydrogen bond with the side chain of K310 in some models of DENV1/DIII (Fig. 38).

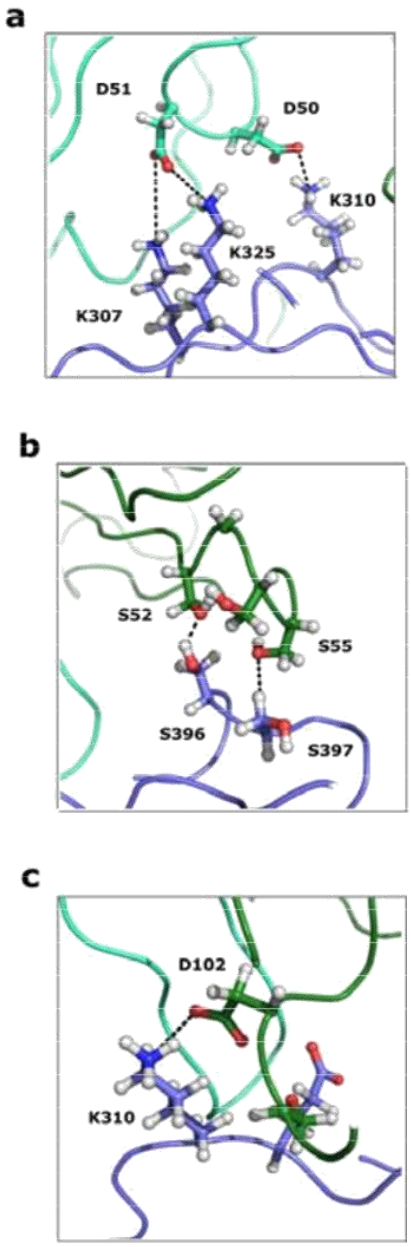

Fig. 38 Atomic details of DV32.6-DENV1/DIII interaction. (a) Hydrogen bond interactions at the core of the Ab/Ag interface between Asp50 and Asp51 of DV32.6 L2 loop and Lys307, Lys310 and Lys325 of DENV1/DIII. (b) Hydrogen bonds between Ser52 and Ser55 of DV32.6 H2 loop and Ser396 and Ser397 of DIII. (c) Ionic interaction between Asp102 in H3 loop and Lys310 of DIII. Nitrogen and oxygen atoms are shown in blue and red,

respectively, whereas hydrogens are in white. Possible hydrogen bonds are indicated by broken lines. DIII is shown in light blue, while the Ab is in green. Model 8 is shown.

To sum up, comparing the structure of the complexes between DV32.6 and DENV1/DIII or DENV4/DIII, our results show that the L2 loop of the antibody, and in particular the two aspartates in position 50 and 51, play an important role in both cases. DV32.6 interacts more extensively with DENV1/DIII rather than DENV4/DIII, having contacts with 5 of its 6 CDR loops. Furthermore, the binding orientation of DV32.6 is different in the two complexes, with implications for the epitope accessibility on the viral surface (Fig. 39)

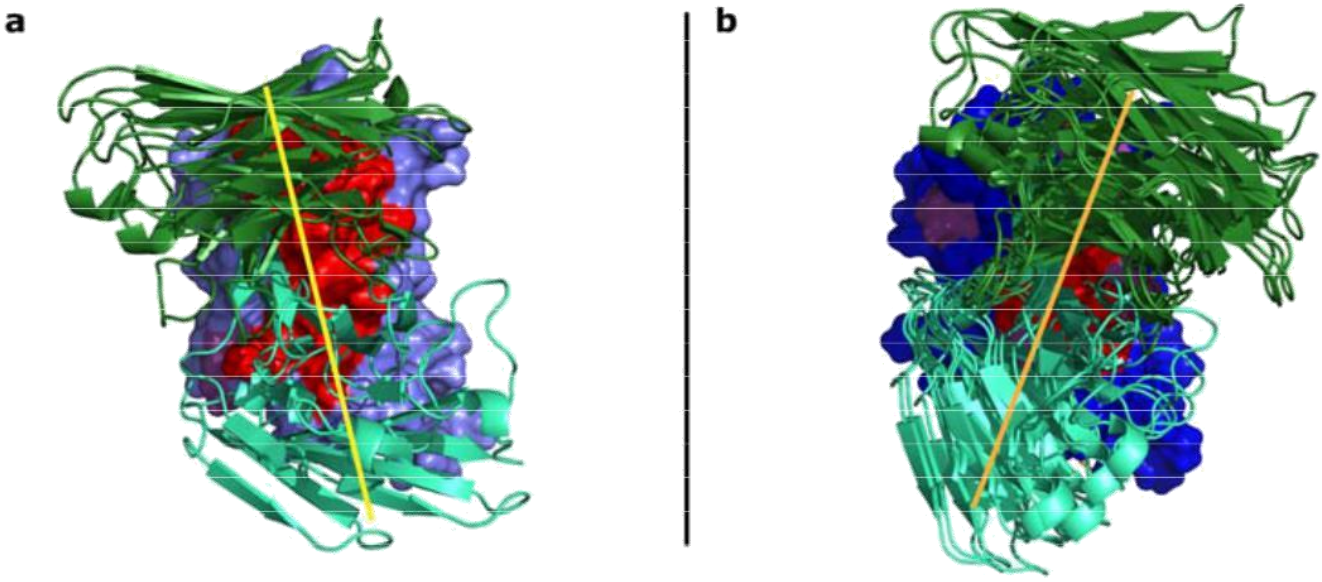

Fig. 39 Different orientation of DV32.6 in complex with DENV1/DIII (a) and DENV4/DIII (b). The DIIIs of the two serotypes are aligned. Only the models that are in agreement with the experimental data are shown. The yellow and

orange lines is the schematic representation of the orientation that the antibody assumes when it interact with DENV1/DIII and with DENV4/DIII.

How DV32.6 interact with DenV serotypes

The step over in the understanding of the interactions between DV32.6 and DenV serotypes was elucidate the structure of DV32.6 in complex with all the remaining DenV serotypes and exploit the differences to rationally design mutated antibodies with i) selectively altered binding specificity and ii) improved ability to neutralize the virus. First of all has been used NMR epitope mapping to define the binding site of DV32.6 on DIII of all four DenV serotypes. Then use this information to filter computational predictions of the antibody/antigen complexes. Analysis of the resulting three dimensional structures proved sufficiently accurate for the rational design of antibody mutants with selectively altered binding specificity or improved neutralization properties.

DV32.6 is part of a panel of human monoclonal antibodies isolated from a donor recovered from infection by DenV2[145]. It binds to DIII of all four DenV serotypes with KD 145±9 nM for

DenV1; 7±0.2 nM for DenV2; 73±16 nM for DenV3; 34±7nM for DenV4 according to SPR (Figure 40).

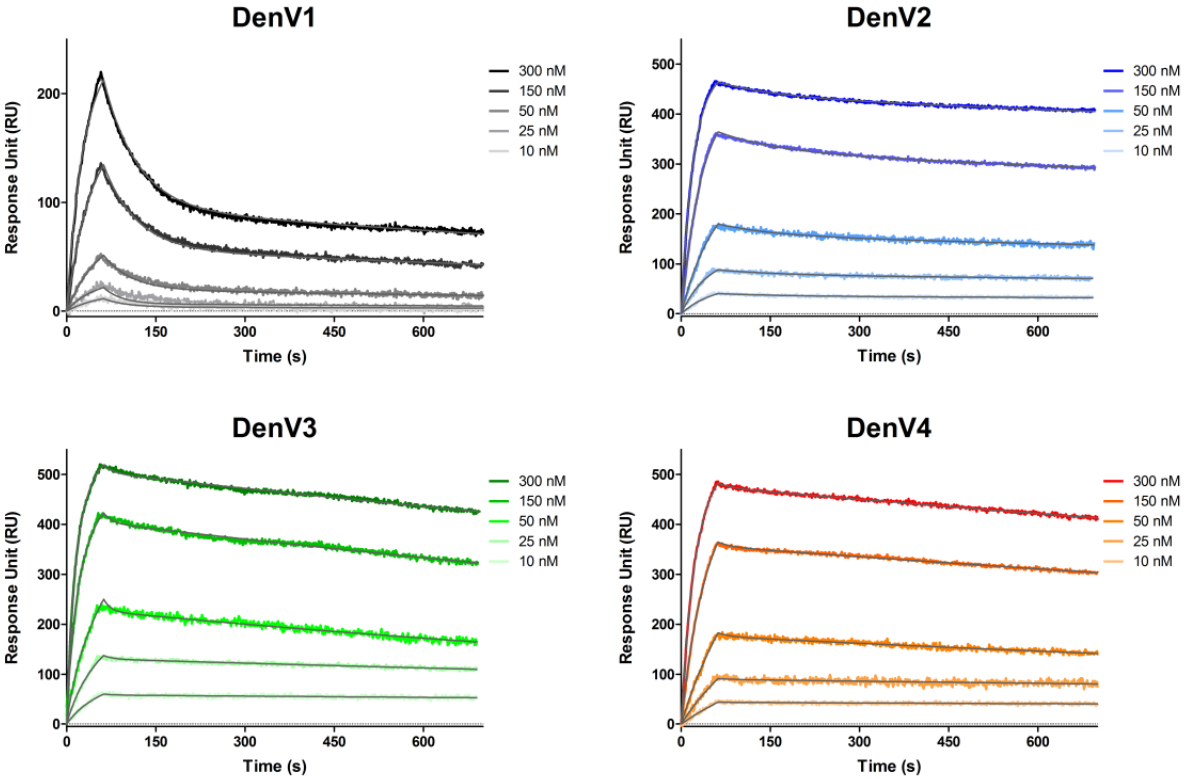

Fig. 40 SPR sensorgrams showing association and dissociation of DV32.6 to DIII of each Dengue serotype. The antibody was immobilized on the sensor surface, followed by injection of DIII at the concentrations indicated in the figure. The line fitted to the experimental data and used to calculate the binding affinities is drawn in gray. KD values are 145±9 nM for DenV1; 7±0.2 nM for DenV2; 73±16 nM for DenV3; 34±7 for DenV4. Despite binding more weakly to DenV1 and DenV3 than DenV4, the antibody neutralizes those serotypes better than DenV4.

The ability of DV32.6 to neutralize the virus was assessed by flow cytometry assays measuring the number of cells infected by DenV vaccine strains in the presence of different amounts of antibody. There is no direct correlation between DIII binding affinity and neutralization: the antibody is more efficient

at neutralizing DenV2, DenV1 and DenV3 despite binding more strongly to DIII of DenV4.

Association and dissociation rates show no obvious correlation to the neutralizing activity, either. The approximate concentration of antibody required to neutralize 50% of the viral activity is 2mg/ ml for DenV2, 3mg/ml for DenV3, 4mg/ml for DenV1 and 74mg/ml for DenV4. Binding assays on isolated DIII allow us to compare the binding affinity of the antibody for its epitope but the natural target for the antibody is the full virus, where the surface of DIII is partially covered by neighbouring protein domains. In contrast to DIII binding, ELISA performed at 37°C on the full virus show that binding to DenV4 is not stronger than to other serotypes.

Solution NMR spectroscopy was used to characterize, at the residue level, the epitope of DV32.6 on DIII from each DenV serotype. In a so-called 15N-HSQC experiment, the position of these signals is sensitive to the local chemical environment, so much so that 15N-HSQCs are considered protein fingerprints. We exploit this property to determine the epitope: when DV32.6 binds to DIII, interface residues experience a different chemical environment and their NMR signal changes as a consequence. By comparing the NMR spectrum of DIII free or in complex with the antibody we can identify which signals change. Knowing the assignments, we can therefore determine the residues affected by antibody binding (Figure 41).

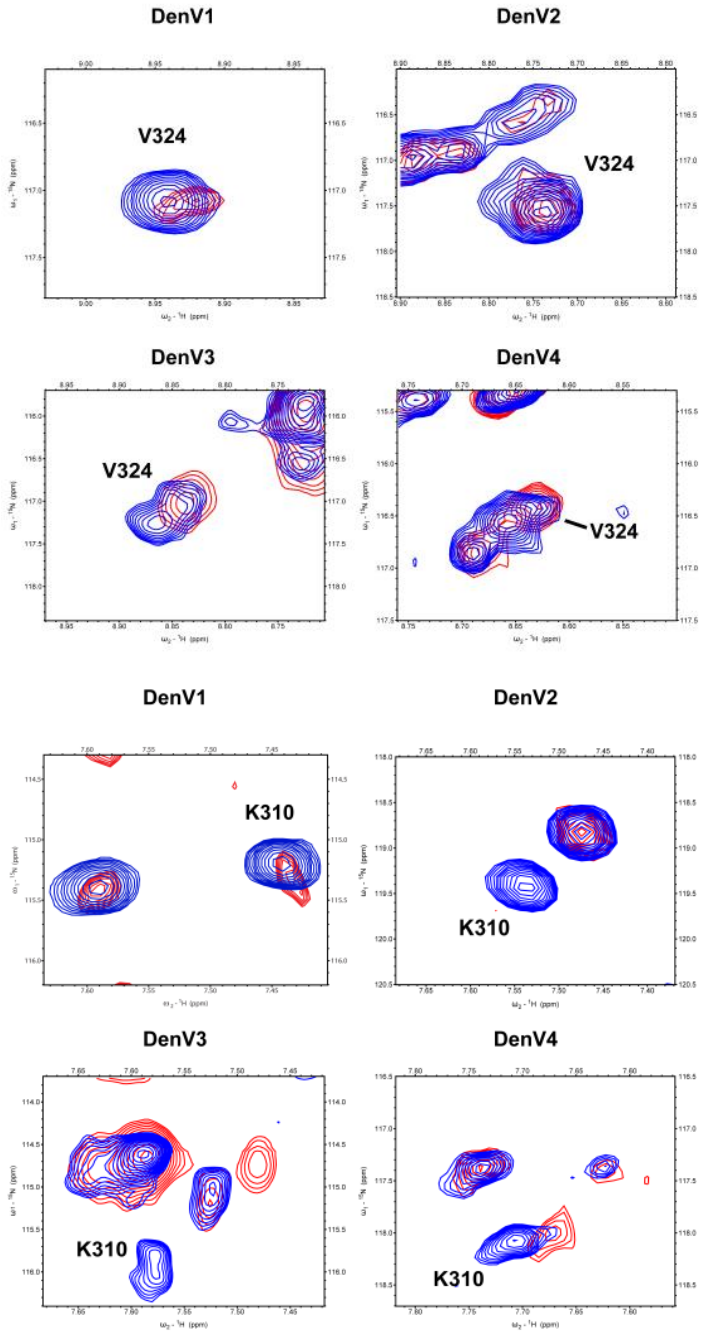

Fig. 41 15N HSQC spectra of DIII of the four Dengue serotypes free (blue) and in complex with antibody DV32.6 (red). Residue V324 is affected by complex formation and shows chemical shift changes in DenV1, DenV3 and DenV4 but not DenV2. Residue K310 shows chemical shift changes upon complex formation in DenV4 and DenV1 (smaller changes). The peak

corresponding to the bound state disappears in DenV2 and DenV3, revealing that the residue is affected by antibody binding.

We assigned DIII of DenV3 according to standard techniques. We utilize a purely qualitative approach: if a signal changes position or disappears, then we know that the residue generating such signal is affected by antibody binding.

The NMR signal of approximately 20% of the surface residues of DIII is perturbed upon binding of DV32.6. No unassigned peak shows chemical shift changes upon complex formation with the exception of one signal that presumably belongs to Q316, in the middle of the epitope. The epitope centers around residues 306–325 and the antibody footprint shows only slight variation amongst serotypes (Figure 42a and 42b).

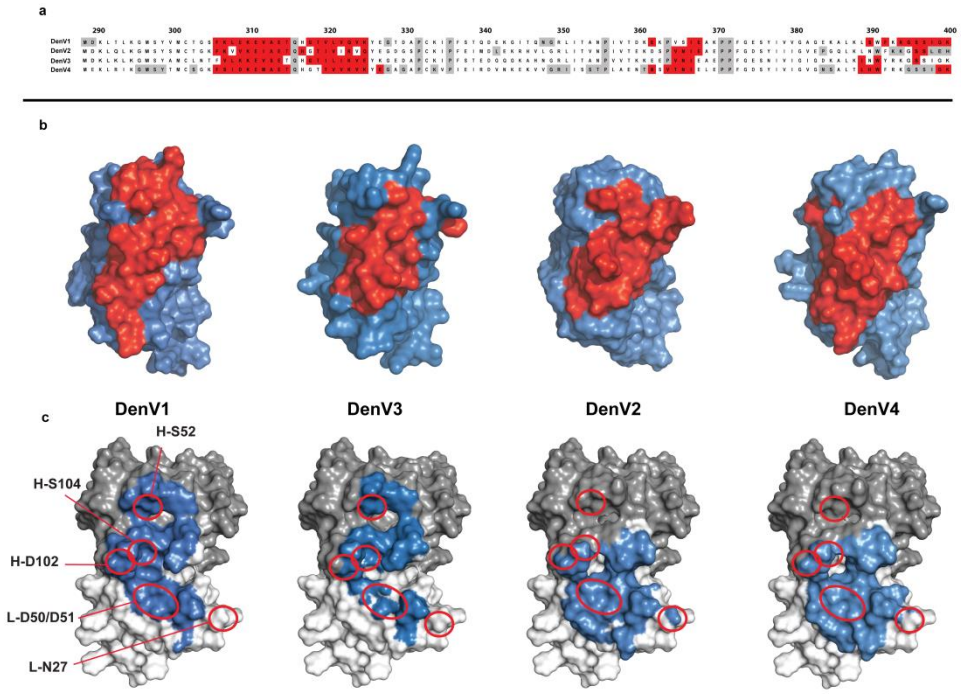

Fig. 42 NMR epitope mapping results (a, b): residues whose NMR signal is affected by antibody binding are indicated in red on the sequence (a) and on the surface representation (b) of DIII of each DenV serotype. Residues for

which no NMR information is available are coloured gray in the sequence. The discontinuous epitope shows slight variations amongst serotypes both in sequence and structure, including some conserved residues and others that are not. The former explain why DV32.6 can bind to all four serotypes, the latter are likely responsible for the different binding and neutralization properties.

Computational docking results (c). Surface representation of the antigen binding site of DV32.6. Antibody residues predicted to interact with DenV are shown in blue; light and heavy chains are in light and dark gray, respectively. Both antibody chains contribute to the binding interface in DenV1 and DenV3 whereas only the light chain and the H3 loop bind to DenV2 and DenV4. Some of the residues mutated to alter the antibody properties in a predictable way (see text) are indicated by red circles and labelled on DenV1.

It includes residues conserved in all serotypes, explaining why DV32.6 binds all four of them, and residues that are not conserved, which are probably responsible for the different binding and neutralization properties. To further characterize the binding interface we thus predicted the three dimensional structure of the complex between DV32.6 and DIII by computational docking, guided and validated by the NMR epitope mapping data.

We predicted the structure of antibody DV32.6 by homology modeling according to the canonical structure method. Then thus generated multiple models of DV32.6[35], differing mainly in the conformation of the H3 and L3 loop and in the relative position of the six antigen binding loops. In particular, the H3 loop extends outwards in some of our models but presents a rather flat

interacting surface in others. The purpose is two-fold: on one hand by using multiple models we increase the chances that at least one of them is accurate; on the other hand the ensemble of conformations may simulate the flexibility available to long protein loops such as the H3 of DV32.6. We then docked each antibody model independently to DIII with the program RosettaDock[148] and validated the results.

According to our NMR validated docking predictions, DV32.6 primarily recognizes DIII residues located on adjacent beta strands, covering between 25 and 28 residues with a buried surface area between 684 $A^{\circ 2}$ (DenV2) and 768 $A^{\circ 2}$ (DenV1). These values are in line with those obtained from x-ray structures of other antibodies against DIII. The predicted interface of antibody DV32.6 is dominated by electrostatic interactions and features several intermolecular hydrogen bonds and salt bridges. All DIII residues predicted to be at the interface by the selected docking solutions are also affected by complex formation in the NMR epitope mapping experiment. K310, conserved in all serotypes, is at the center of the interface and involved in multiple intermolecular contacts. The epitope has several non-conserved residues, which are probably responsible for the different binding and neutralization properties of DV32.6. Indeed, residues 307, 309, 323, 325 and 327 differ among serotypes and appear to be involved in antibody binding. Residue 309, for instance, is Glu, Val, Lys or Asp in DenV1 to DenV4, respectively, and it may not be surprising that it has a different effect on the antibody partner. Perhaps surprisingly, some conserved DIII residues appear to have different antibody partners in different serotypes:

Analyzing the predicted antibody interface (Figure 42c) shows differences that could be exploited to selectively alter the binding properties: the light chain (L1, L2, L3) and H3 antigen binding loops interact with every serotype; the H2 loop interacts only with DenV1 and DenV3 and the H1 loop has no contact with any serotype. Although the total interface area is similar, the light and heavy chains equally contribute to it in DenV1 and DenV3, whereas 80% of the interface is formed by the light chain in DenV2 and DenV4

Rational Antibody Engineering to Selectivity Alter its Binding Properties

By analyzing the three-dimensional models of the antibody/DIII complexes we designed several antibody mutants with the intent of further verifying our computational predictions, altering the binding properties of the antibody and ultimately improving its ability to neutralize the virus. As a first test we aimed to modify the antibody without affecting its binding to DIII, proving that we can identify and avoid critical residues. Antibody sequence analysis can easily predict which residues belong to antigen binding loops and may, therefore, interact with the antigen. Our models go a step further and can identify which of these residues are not directly involved in intermolecular contacts and can thus be mutated without adverse consequences. We thus selected and mutated a subset of such residues in each of the three heavy chain antigen binding loops. As predicted, all the following mutations in the antigen binding loops

didn’t alter the antibody binding properties: H-S104A (Figure 3); H-T31A; H-S54Q; H-S103V; H-T106A.

To further validate our models we then aimed to abolish antibody binding to all serotypes. Residues L-D50 and L-D51 are predicted to be at the center of the antibody/DenV interface, forming an intermolecular network of hydrogen bonds and salt bridges in all serotypes. If our models are correct, disrupting this network should have a profound effect on the interface. Indeed, the double mutant L-D50A/D51A completely abolishes antibody binding (Figure 3). The same result is obtained when mutating the nearby H-S105D, since the negative charge introduced interferes with the negatively charged sidechain of L-D50. Having shown that our approach can identify critical interface residues, we tackled the much more difficult task of improving the antibody properties. Generally speaking, increasing antibody selectivity is a useful exercise to eliminate cross-reactivity with undesired antigens or to design bio-recognition elements for specific antigen subtypes. As a proof of concept, we altered the interaction between DV32.6 and DenV, obtaining an antibody mutant specific for DenV2 and another that binds only DenV2 and DenV4. According to our models, residues H-D102/S103, belonging to the H3 antigen binding loop, point away from the DenV2 antigen. Therefore, mutating them should have no effect on this serotype. They are, instead, in close proximity of DenV4 and DenV3. Finally, H-D102 is predicted to form an intermolecular salt bridge in DenV1. In agreement with this prediction, the HD102A/S103A mutant binds DenV2 like the wild-type antibody but it has severely reduced binding to all other serotypes (Figure 3).

Since the antibody H2 loop is predicted to interact with DenV1 and DenV3 but not DenV2 and DenV4, we designed the H-S52A mutant to prevent antibody binding to DenV1 and DenV3 while leaving the other two serotypes unaltered (Figure 3). This mutant agrees with the computational prediction that the antibody uses different binding modes to interact with the serotypes as illustrated in Figure 2c.

Rational Antibody Engineering to Improve the Neutralization Properties

The aforementioned mutations are a testimony to our ability to rationally alter the antibody binding properties and specificity. Improving the ability of therapeutic products to protect from infection is, however, the main goal of antibody engineering. As a proof of concept, analysis of the predicted antibody/antigen interfaces allowed us to design a mutated antibody 40 times more efficient than the wild-type in neutralizing DenV1 and another mutant more efficient against all serotypes, albeit to a lesser extent (Figure 43).

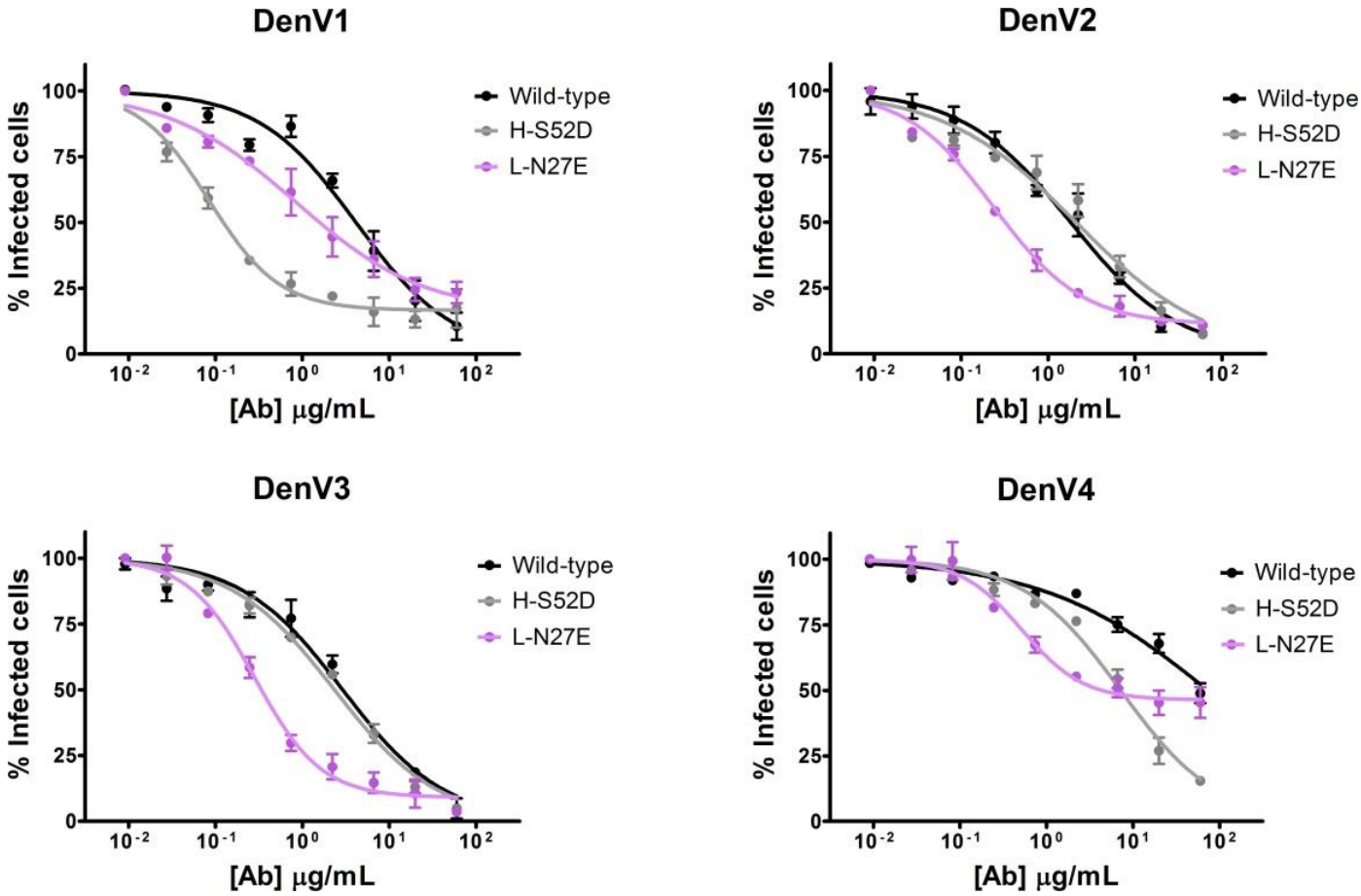

Fig. 43 We designed two antibody mutants with the intent of improving its neutralization properties. H-S52D (gray) neutralizes DenV1 40 times more efficiently than the wild-type (black) and L-N27E (violet) is better than the wild-type in all serotypes, albeit to a lesser extent. Viral neutralization assays are shown; the amount of infected cells (y axis) decreases at increasing antibody concentration (x axis). In comparison to the wild-type, a lower concentration of mutants is required to neutralize the same amount of virus.

Our models show that H-S52 is in close proximity to a positively charged lysine sidechain in DenV1 (Figure 5).

Introducing a nearby negative charge should favor the formation of intermolecular salt bridges, possibly resulting in improved neutralization properties. Indeed, the H-S52D antibody is 40 times more efficient in neutralizing DenV1; the estimated antibody concentration required to achieve 50% neutralization (EC50) is 0.1 mg/ml for HS52D and 4.1mg/ml for wild-type DV32.6 (Figure 43).

SPR indicates that the k_{off} is identical to the wild-type while the k_{on} improves from $0.95\pm0.01x10^{-4}$ to $4.8\pm0.4x10^{-4}$ $nM^{-1}s^{-1}$ resulting in a KD of 28±2 nM versus 145±9 nM. DenV4, where H-S52D is also a slightly better neutralizer than the wild-type, has a similar improvement in k_{on} ($1.5\pm0.1x10^{-4}$ for the wild-type,$5.3\pm1.0x10^{-4}$ $nM^{-1}s^{-}$ for H-S52D) and the k_{off} improves as well (from $5.1\pm1.3x10^{-3}$ to $0.47\pm0.1x10^{-3}$ s^{-1}) resulting in KD of 0.9±0.04 nM for the mutant versus 34±7 nM for the original antibody. L-N27, instead, is predicted to be close to a positively charged region in all serotypes (Figure 44) so introducing a negative charge should favor the interaction with all of them.

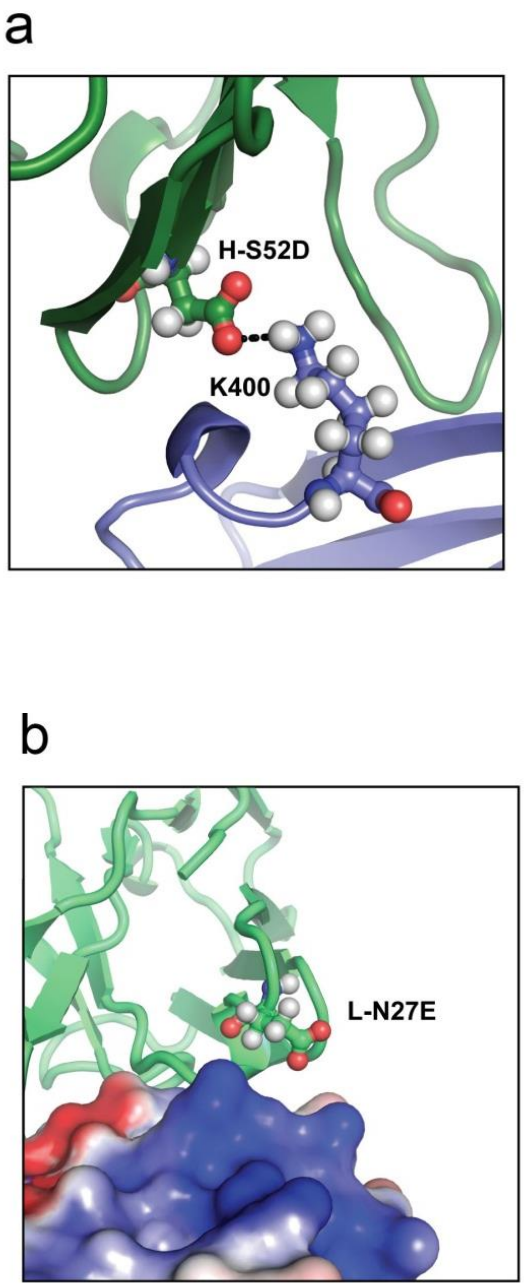

Fig. 44 a) H-S52 is close to the positive sidechain of K400 in DenV1. Mutating it to the negatively charged H-S52D favours the interaction and this mutant is 40 fold more effective than the wild type at neutralizing DenV1. The antibody is shown in green and Denv1 in blue.

b) DV32.6 is shown as green cartoon over the electrostatic surface of DenV3; positively charged surfaces are in blue and negatively charged surfaces in red.

L-N27 is close to a positive patch conserved in all serotypes. Introducing a negative charge (L-N27E) favours the interaction, resulting in improved neutralization properties. The L-N27E substitution was preferred to L-N27D since the longer sidechain was thought to bring the charged antibody moiety closer to DIII.

An N>E rather than N>D mutation was introduced because a longer side chain would get closer to the antigen without creating clashes, according to the models. Indeed, the L-N27E mutant has improved neutralization properties for all serotypes (Figure 43).

The L-N27E mutant is 3 times more effective on DenV1, 6 on DenV2, 9 on DenV3 and 17 on DenV4 (ratios between the EC50 for wild-type and mutant antibody). KD is 5 times stronger for DenV1, 23 times stronger for DenV2, 4 times for DenV3 and 38 times for DenV4. The mutant has improved K_{on} in all serotypes whereas the K_{off} is equal to the wild type for DenV1 and DenV3. Unfortunately, the effects of the H-S52D and L-N27E mutations are not additive; the double mutant shows better neutralization only for DenV3. Apparently, the interaction cannot tolerate the simultaneous introduction of two negative charges on the other serotypes

Conclusions

The test case that has been performed using a known X-ray structure of a complex between DIII of DENV2 and the antibody 1A1D demonstrates that docking is able to give us the correct solution, but relying only on the scoring function makes it impossible to discriminate the right solution amongst the thousands of wrong ones.

Here we utilized NMR chemical shift mapping to identify the epitope of an antibody isolate from a Dengue patient, called DV32.6, on DIII of DENV4.DV32.6 binds the whole virus, the full E protein dimer and DIII alone in all four dengue serotypes but it can not neutralize DENV4 and only weakly neutralizes the other serotypes, showing also a significant ADE activity.

According to our NMR epitope mapping experiment, applied for the first time to the characterization of a full size human mAb, there are twenty-six residues of DENV4/DIII that show significant chemical shift perturbation upon formation of the complex, defining the epitope region. We then performed the docking simulations to obtain information about the antibody residues implicated in the binding and about the orientation of the two partners of interaction. Models with a good scores are analyzed to check their agreement to the experimental data both by visual inspection, initially, and also at a finer and more quantitative level by using two indicators that we developed specifically for the task. The validity of such indicators was tested on a TCR-pMHC complex in which NMR mapping was later proven accurate by an x-ray structure[24].

The docking simulation was performed for each of the eleven different antibody structures independently obtaining an ensemble of decoys. These are filtered according to the above-mentioned criteria and only the models that are in agreement with the experimental data are kept. This approach allows us to take in account the conformational flexibility of the CDR loops of the antibody, especially of H3, into the docking search, considerably improving the accuracy of the result.

The same procedure was performed for the DV32.6-DENV1 interaction. According to the chemical shift mapping experiment, the epitope residues are almost the same of DENV4 but comparing the structures of the two complexes shows a different binding orientation of the antibody. Analyzing the results in the context of the full E protein dimer, which forms the viral surface, the epitope on DENV1 is more accessible than that of DENV4. Considering that, according to our neutralization assays, DV32.6 is capable of neutralizing DENV1 but not DENV4 despite a stronger binding to DENV4-DIII, we can speculate that this is due to the partial occlusion of the DENV4 epitope by the other domains of the E protein.

Moreover strategies for optimizing and improving antibody properties are highly desirable, either to increase their efficacy or to alter their binding specificity. Generally speaking, abolishing or altering antibody binding to undesired antigens might confer selectivity to an otherwise broadly reactive antibody, with practical uses in avoiding cross-reactivity or designing specific biomarkers. Here we move further ahead by showing that computational models of antibody/antigen complexes allowed us to rationally design

antibody mutants that improve its properties by i) disrupting binding only to selected serotypes and ii) improving virus neutralization up to 40 fold. Antibody DV32.6, isolated from the serum of a human donor recovered from DenV2 infection, can efficiently neutralize three of the four DenV serotypes but binds to all four in the so-called DIII, the most variable domain of the protein forming the viral surface and the target of many potent antibodies described so far. We identified the epitope of DV32.6 on each serotype at the residue level with solution NMR spectroscopy and used this information to guide and validate computational docking simulations yielding three-dimensional models of the antibody/antigen complexes. Visual analysis of these NMR validated computational structures allowed us to design a total of 22 antibody mutants, only 4 of which did not have the predicted effect. With this approach we produced DV32.6 antibody mutants that could a) leave binding unaltered, proving that we can identify residues that are not critical for interaction despite belonging to the antigen binding loops or b) abolish binding to all serotypes. More importantly, whereas wild-type DV32.6 binds to all four DenV serotypes, we could design antibody mutants binding only to DenV2 or to DenV2 and DenV4. Increasing antibody specificity is valuable to eliminate unwanted cross-reactivity or to design bio-recognition elements. If an antibody has therapeutic purposes, however, one would seek to improve its neutralization properties, which would have a beneficial effect on dosage and therapy.

In conclusion, we demonstrated that computational docking is a fast and accurate tool to obtain the structure of antibody/antigen complexes, provided that it is validated by rapidly obtained experimental data. The use of NMR

epitope mapping, a fast and reliable technique to identify antibodies epitopes, noticeably improves the docking accuracy.

General conclusion

Solution NMR spectroscopy is ideally suited to the characterization of intermolecular interfaces, including antibody/antigen complexes. NMR chemical shift mapping provides a detailed description of the binding footprint of one molecule over another.

The results of NMR epitope mapping were used to greatly increase the accuracy of computational docking, providing a three-dimensional atomic structure of the protein/protein and antibody/antigen complex. Roughly speaking, information on the NMR-derived epitope is either used as a constraint to guide the structure calculation algorithm[30,150,151], possibly with other information that can help orientate the components of a multimolecular complex[152], or at the end of the computational simulation to filter out models that do not agree with the experimental epitope[34,35]. In the antibody/antigen interaction the resulting structures were sufficiently accurate to allow rational modifications of an antibody, altering its selectivity and increasing its viral neutralization properties by up to 40 folds[36]. On the other hand, the NMR mapping used to drive computational docking was useful to hypnotize the structural conformation of CXCL12/HMGB1 complex in the cellular response triggering. Although Xray is obviously better at structure determination, NMR mapping offers comparable results in a fraction of the time for epitope characterization.

Once NMR assignments are available for the antigen or general proteins, several different partners can be rapidly mapped, typically requiring few days.

NMR epitope mapping also has a clear advantage over peptide mapping and site-directed mutagenesis, because it provides a complete description of the antigen residues contacted by the antibody which is very difficult to obtain by these other techniques. The disadvantages of NMR epitope mapping include the fact that the antigen or in this case chemokines needs to be labeled with NMR active nuclei (typically 15N). It is thus necessary to express the antigen in Escherichia coli (more seldom yeast), which is not always possible. Furthermore, large antigens present a challenge for NMR due to both spectral overlap and signal broadening at increased molecular weight leading to loss of sensitivity. Selective labeling of specific antigen residues, advanced acquisition techniques such as TROSY and the use of 15N-HSQCs, probably the simplest and most sensitive multidimensional NMR experiment, somehow alleviate these limitations, which nonetheless remain significant in some cases. Finally, NMR assignment of the antigen, that is, knowing which NMR signal corresponds to which antigen residue, is necessary. Assignment can be a trivial process of a few days or a difficult task requiring several months and complex strategies for selective labeling.

NMR epitope mapping is a rapid and accurate tool for the characterization of antibody/antigen or protein/protein complexes. So far, there are relatively few examples of using high-resolution NMR spectroscopy to determine antibody epitopes, and most involve small peptides rather than full protein antigens[35,36,37,38,39,42,153], but NMR epitope mapping is a valuable tool for the detailed structural characterization.

Material and methods

Protein production and purification

^{15}N-labeled CXCL12 was obtained by growing *Escherichia coli* Rosetta cells transfected with pET30 vector, in which the CXCL12 sequence was cloned, in M9 minimal media containing ^{15}N- ammonium chloride., The cells were induced at OD_{600}=0.6 and harvested after 4 hours. The pellet was put for at least 1 hour at -20 °C. The cells were thawed and resuspended in 50 mM TrisHCl pH8, 100 mM NaCl; 1 mM EDTA 5 mM DTT and 0.1% Triton X100. After sonication and centrifugation, the pellet was repeatedly washed and centrifuged with 50 mM TrisHCl pH8, 100 mM NaCl; 1 mM EDTA 5 mM DTT and 0.1% Triton X100. and subsequently solubilized in 50 mM MES pH 6.5 and 6 M guanidinium HCl overnight at 37 °C. Afterwards CXCL12 was refolded by dilution in 50 mM TrisHCl pH8, 50 mM NaCl, 0.1 mM reduced glutathione and 0.1 mM oxidized glutathione, overnight at 4°C. CXCL12 was then loaded on SP Cation exchange column and finally passed in Size exclusion column Superdex75 pre-equilibrated with20 mM NaP pH6 ,20 mM NaCl.

^{15}N-labeled BoxB was obtained by growing *Escherichia coli* Rosetta cells transfected with pET30 vector in which the BoxB sequence was cloned, in M9 minimal media containing ^{15}N- ammonium chloride. The cells were induced at OD_{600}=0.8-0.9 and harvested after overnight growing. The pellet was put for at least 1 hour at -20 °C. The cells were thawed and resuspended in 20 mM TrisHCl pH8, 10 mM imidazole, 2 mM 2-β-mercaptoethanol, 150 mM NaCl. After sonication and centrifugation, the supernatant was loaded on Hitrap

Chelating HP column; after the loading 2 washing steps with increasing concentration of imidazole were performed and BoxB was eluted with 300 mM of imidazole. The protein was then dialyzed to remove the imidazole and passed on the size exclusion Superdex75 column pre-equilibrated in 20 mM NaP pH6 and 20 mM NaCl.

HMGB1 was obtained by growing *Escherichia coli* Rosetta cells transfected with pET30 vector, in which the HMGB1 sequence was cloned, in M9 minimal media containing ^{15}N- ammonium chloride., The cells were induced at OD_{600}=0.7 and harvested after 3h. The pellet was put for at least 1 hour at -20 °C. The cells were thawed and resuspended in 20 mM TrisHCl pH8, 10 mM imidazole, 2 mM 2-β-mercaptoethanol, 150 mM NaCl, 0.2% triton X100. After sonication and centrifugation, the supernatant was loaded on Hitrap Chelating HP column ; after the loading3 washing steps with increasing concentration of imidazole were performed and HMGB1 was eluted with 300 mM of imidazole. The protein was then dialyzed to remove the imidazole and passed on the size exclusion Superdex75 column pre-equilibrated with 20mM TrisHCl pH8 and 20 mM NaCl and 2 mM 2-β-mercaptoethanol.

HMGB1 purification; schematic practical protocol

Day 1

Transformation of Rosetta cells

Used 150 μL of competent cells for transformation.

De-ice competent cells.

Add DNA 50 ng HMGB1

Incubate at 4 ˚C for 30 min.

Thermal shock at 42 ˚C for 37 sec.

Incubate for some minutes on ice.

Add 600 μL of LB and incubate at 37 ˚C for 1 h.

Plate on LB + KAN + CHO (~150 μL)

Incubate at 37 ˚C ON.

Day 2

Starter cultures

Starting from single colony pre-colture in 50 mL LB + KAN (30 μg/mL) + CHO (34 μg/mL).

Incubate at 37 ˚C ON in agitation.

Day 3

Growing cultures

Used the pre-colture to inoculate 1 L of LB + KAN + CHO; starting OD_{600} about 0.1

The colture was grown at 37 ˚C, rotation 220 rpm.

Induction

OD_{600} : ~0.7

Induced with 0.5 mM IPTG (growing always at 37 ˚C).

Harvest

After 3 h the cells were collected by centrifugation (6000 rpm for 20 min at 4 ˚C).

The cells were then store at -20 ˚C.

Day 4

Protein extraction

Resuspend the pellet into the appropriate volume (~ 2 mL for g of cells) of LYSIS BUFFER (20 mM TrisHCl pH 8; 150 mM NaCl; 10 mM Imidazole; 2 β-SH; 0.2% Triton X100).

Sonication: 30 sec of pulsing/30 sec of pause for 4-5 times
Centrifuge at 18000 rpm for 30 min at 4 °C (**rotor ss-34**); keep the pellet (keep the SN for SDS Page)

Hitrap Chelating Purification (Nickel column)

Filter the SN and load on Hitrap Chelating HP column (GE Healthcare 5 mL volume) pre-equilibrated in lysis buffer.
After loading wash as follow:

- Wash 1: lysis buffer – (20 mM TrisHCl pH 8; 150 mM NaCl; 10 mM Imidazole; 2 β-SH)

- Wash 2: wash 2 buffer (20 mM TrisHCl pH 8; 1M NaCl; 10 mM Imidazole; 2 β-SH)
- Wash 3: wash 3 buffer (20 mM TrisHCl pH 8; 150 mM NaCl; 30 mM Imidazole; 2 β-SH)

  After the washing steps elute the protein with elution buffer (20 mM TrisHCl pH 8; 150 mM NaCl; 300 mM Imidazole; 2 β-SH)

Collected the fractions of the last peak (after 300 mM imidazole)

Put in dialysis ON (1 L + 1 L) dialysis buffer 20 mM TrisHCl pH 8; 20 mM NaCl; 2 β-SH)

Day 5

The dialyzed sample was loaded on SE75 HiLoad pre-equilibrated with 20 mM TrisHCl pH 8; 20 mM NaCl; 2 β-SH.

According to the column specification, the peak corresponding to HMGB1 molecular weight is around 60 mL.

LPS purification

To remove LPS load your purify HMGB1 on an Anion Exchange (Q) column.

Loading buffer: 20 mM NaP pH7.2; 20 mM NaCl

Elution buffer: PBS pH 7.2+1.5M NaCl

Put in dialysis ON (1 L + 1 L dialysis buffer PBS pH7.2)

BoxA and BoxB purification; schematic practical protocol

Transformation of Rosetta cells

Used 150 µL of competent cells for transformation.

De-ice competent cells.

Add DNA 50-100 ng

Incubate at 4 ˚C for 30 min.

Thermal shock at 42 ˚C for 37 sec.

Incubate for some minutes on ice.

Add 600 µL of LB and incubate at 37 ˚C for 1 h.

Centrifuge for 10 min at 4000 rpm

Discard about 600 µL of supernatant (SN), resuspend the pellet in the remaining LB and plate on LB + KAN + CHO (~50-60 µL)

Incubate at 37 °C ON.

Day 2

Starter cultures

Starting from single colony 2 pre-colture in 50 mL M9 + KAN (30 µg/mL) + CHO (34 µg/mL).

Incubate at 37 ˚C ON in agitation.

Day 3

Growing cultures

Used the pre-colture to inoculate 2 flasks of 2 L containing 500 mL of M9 + KAN + CHO.

All the cultures were grown at 37 ˚C, rotation 220 rpm.

Induction

OD_{600} : 0.8-0.9

Induced with 1 mM IPTG (**now growing at 25 ˚C**).

Day 4

Harvest

After ON the cells were collected by centrifugation (6000 rpm for 20 min at 4 ˚C).

Amount of cells: about 7 gr

Store the cells at -20 ˚C

Day 5

Protein extraction

Resuspend the pellet (~7g) into 20 mL of LYSIS BUFFER

Lysis buffer
20 mM Tris-HCl pH 8
10 mM imidazole pH 8
2 mM β-Mercapto
150 mM of NaCl
0.2% Triton
DNAse

SONICATION: 30 sec of pulsing/30 sec of pause for 4-5 times

Centrifuge at 40000 rpm for 30 min at 4 °C or 18000 for 1h, 4°C;

keep the SN and filter it.

Hitrap Chelating purification

Column: Hitrap Chelating HP 5 mL (GE Healthcare)
Load on the 5 mL column maximum 60 mg of protein

Equilibrate the column with H_2O (20 mL)

Load 0.1 M $NiSO_4$ (10 mL)

Equilibrate the column with ATTACH BUFFER (until the UV is stable).

Load the sample and collect the flow-through.

Wait until the UV is stable and then wash the column with WASH 1 and collect the flow-through.

Wait until the UV is stable and then wash the column with WASH 2 and collect the flow-through.

Wait until the UV is stable and then elute with ELUTION BUFFER and collect fractions.

Measure all the fractions and the flow-through with Nanodrop.

Put together for the refolding the fractions containing the protein

Attach Buffer
20 mM Tris-HCl pH 8
10 mM imidazole pH 8
2 mM 2-β-mecaptoethanol
150 mM NaCl

Wash 1
20 mM Tris-HCl pH 8
10 mM imidazole pH 8
2 mM 2-β-mecaptoethanol
1 M NaCl

Wash 2
20 mM Tris-HCl pH 8
30 mM imidazole pH 8
2 mM 2-β-mecaptoethanol
150 mM NaCl

Elution buffer
20 mM Tris-HCl pH 8
300 mM imidazole pH 8
2 mM 2-β-mecaptoethanol

150 mM NaCl

The fractions pulled (i.e. B7/B6 ;3 ml+3 ml) were diluted until 20 ml with the Elution Buffer **without imidazole** and after concentrate until 3 ml to make Size Exclusion chromatography.

Size exclusion chromatography

Column: Hiload Superdex75 (GE Healthcare); column volume about 125 mL
Load on the column maximum 3 mL of protein
Buffer: 20 mM NaP pH= 6 + 20 mM NaCl

DenV DIII purification; schematic practical protocol

- **PROTEIN EXPRESSION**

- Transform Rosetta cells (DE3) with DIII plasmid and perform eat-shock protocol (30 sec @ 42 °C and 2 min in ice. Add SOC media and leave @ 37 °C for 1 h.
- Place the cells in LB plate + Amp + CHA and grow O.N.
- Grow cells @ 37 °C until the O.D.600 is around 0.6-0.8 and then induce with 0.5 mM IPTG.
- After induction grow @ 37°C for 4 hs, then collect the cells with a SORVALL centrifuge (6000 rpm for 20 min)

- **PROTEIN EXTRACTION**

(set the centrifuge @ 4°C!!)

(tube in ice!!!)

- resuspend the pellet (500mL) into 30mL of 20mM NaPhos + 150mM NaCl pH7.5
- SONICATION: 1’ of pulsing/1.5’ of pause for 6 times
- Centrifuge @ 18000 rpm for 30’ @ 4°C
- WASHING: resuspend the pellet into 25-30 mL of washing buffer (20mM Na phos pH7.5 + 150mM NaCl + 1M urea + 1% Triton) and centrifuge @ 16000 rpm for 30’
- Resuspend the pellet into 35 mL of solubilization buffer (20mM Na Phos + 8M urea + 1M NaCl)

- Stir @ 37°C O.N.

- **REMOVE DNA**

- Add 0.2% of PEI and leave @ R.T. for 20’
- Centrifuge @ 24000 rpm for 40’ with ultracentrifuge
- Keep the supernatant

- **REMOVE PEI WITH AMMONIUM SULFATE (AS)**

- Add 65% of A.S. and leave 1h
- Centrifuge (17000 rpm for 3’ @ 4°C) and keep the pellet
- Resuspend the pellet into 8M urea + 20mM Na phos

- **REFOLDING**

Refolding by dilution in 20 mM NaP pH 7.5; 100 mM NaCl; 200 mM arginine; 1 mM Glutathione Red; 0.1 mM Glutathione Ox (**final pH about 10**).

Recommended final protein concentration: 0.1 mg/mL

Keep the sample Overnight at 4 °C without stirring.

Size Exclusion Chromatography

Filter the samples, then concentrate with Vivaspin Sartorius concentrators; initial speed 2300g then switch to 1600g at smaller volumes; temperature 8 °C.

Load the concentrated sample on SEC in 20 mM NaP pH= 6 + 50 mM NaCl

NMR epitope mapping

NMR chemical shift mapping were conducted in order to identify the CXCL12 residues affected by the binding with the full HMGB1 and also with its domains separately, BoxA and BoxB. The NMR experiment was also performed on the complex between ^{15}N labeled BoxB and CXCL12.

15N-HSQC experiments were performed on a Bruker 700-MHz spectrometer with CryoProbe. Acquisition time varied between 30 min (for free CXCL12 and BoxB) and 16 h (for the complexes). An HSQC spectrum was recorded on the free, labeled component; a second HSQC was recorded after addition of the unlabeled partner (1:1 ration for the CXCL12/BoxB or BoxA complexes; 2:1 for CXCL12/HMGB1). Backbone assignments were obtained by comparison to previously published data (Biological Magnetic Resonance Bank accession numbers 15148, 15149, and 16145).

Residues whose NMR signal changes position upon complex formation were identified by visual comparison of free and bound spectra of CXCL12 or BoxB in complex with partners. For quantitative analysis, the amount of chemical shift change upon complex formation was calculated for each residues according to the formula $\sqrt{(\Delta N \cdot 0.2)^2 + \Delta H^2}$, where ΔN and ΔH are the chemical shift difference between free and bound spectrum in the nitrogen and proton dimension.

Surface Plasmon Resonance binding assays

Surface Plasmon Resonance (SPR) is a powerful, label-free technique to measure biomolecular interactions in real-time. One of the two partners under study is immobilized to the sensor surface, usually covered with a gold layer,

the other molecule is free in solution and flowed over the surface. Association and dissociation parameters are measured and displayed in a graph called sensorgram (figure 28). Subsequent data analysis allows:

- confirming the binding of two or more partners to each other
- determining the binding affinity of the interactions
- measuring the association and dissociation rates.

SPR-based instruments use an optical method to measure the refractive index near (within ~300 nm) the sensor surface, usually a gold surface. Gold is used because it gives a SPR signal at convenient combinations of reflectance angle and wavelength and, in addition, is chemically inert to solutions and solutes typically used in biochemical contexts. The binding partner (analyte) is injected under continuous flow in aqueous solution (sample buffer) through the flow cell. As the analyte binds to the ligand the accumulation of protein on the surface results in an increase in the refractive index, altering the reflection of the laser light beam sent on the chip. This change in refractive index is measured in real time, and the result plotted as response or resonance units (RUs) versus time (sensorgram).

The sensor chip of the Bio-Rad Proteon-XPR36, used in our experiments, is formed by six parallel channels thus up to six ligand can be immobilized in a single injection step (figure 45).

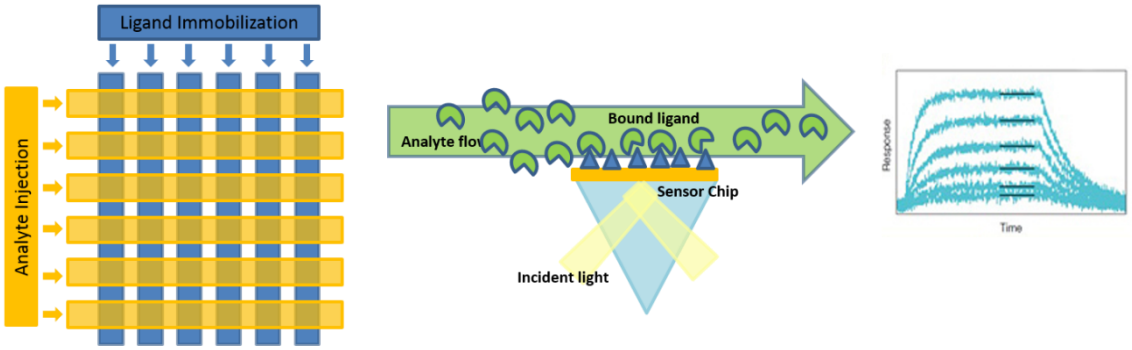

Figure 45: Representation of Bio-Rad Proteon-XPR36 SPR chip surface, the ligand immobilization and analyte injection create a 6x6 ligand-analyte interaction array, and schematic view of the physical process leading to the sensorgram graph.

Immobilization of the ligand to the sensor surface must, of course, happen without disrupting its activity. Several chemical strategies for immobilization are available, the most common for biomolecules involving covalent coupling between carboxyl groups on the sensor surface and amines in the protein ligand.

Amine coupling requires two main steps. Firstly EDAC (1-Ethyl-3-[3-dimethylaminopropyl] carbodimide hydrochloride) reacts with the carboxyl groups on the chip surface forming an amine-reactive intermediate; in the second step the addition of Sulfo-NHS stabilizes the intermediate by converting it to an amine-reactive Sulfo-NHS ester that can form a stable amide-bond to amine groups in the ligand protein (typically lysines). The immobilization is concluded with a blocking step, where a high concentration of ethanolamine is injected to block the remaining activated carboxymethyl groups and prevent further covalent interaction between the surfaces and protein ligands subsequently added over the chip.

Immobilization is followed by the interaction phase. The analyte diluted in the sample buffer is injected into another set of six parallel channels, orthogonal to the six ligand channels, to create a 6x6 ligand-analyte interaction array. Six set of six sensorgram are rapidly generated in a single analyte injection step. Detailed kinetic data can be obtained from these sensorgrams on the interaction of up to six analyte with up to six different ligand.

All the SPR experiments were conducted at 25 °C with Bio-Rad Proteon-XPR36.

The first experiment has been performed to set up the best binding condition for CXCL12/HMGB1. In three of the six vertical channels HMGB1 250 nm in oxidated and reduced states were immobilized, in two of these CXCL12 was immobilized and the last one was empty in order to appreciate the unspecific binding effect. The blocking solution was used in order to block the remaining activated carboxymethyl groups. Afterword the SPR system injected the analyte using the six horizontal channels. In the horizontal direction 5 concentration of CXCL12 have been injected (750, 500, 250 and 100 nM) diluted in three different Buffers with different pH. Phoshate Buffer pH6, Na2HPO4 Buffer pH 11 and NaAc pH 4.5 promoting CXCL12 in monomeric conformation, the sixth channel was empty. In the sixth channel is possible appreciate the large contribution coming from the unspecific binding, that doesn't give the possibility to estimate the binding constant. Otherwise the experiments showed the concentration dependent binding process, indeed the sensorgram signal increase with the increasing of CXCL12 concentration.

In the second experiment, glycyrrhizin was flowed over the surface at concentrations of 0, 500, or 5,000 nM, followed by the addition of 500 nM CXCL12. CXCL12 binding to HMGB1 was detected in all cases, but the signal intensity decreased at increasing concentrations of glycyrrhizin. The experiment was conducted at pH 6.0 in 20 mM Na phosphate buffer and 20 mM NaCl. Several reference channels were used, either flowing CXCL12 and

glycyrrhizin over a sensor surface in the absence of HMGB1 or flowing buffer without CXCL12 or glycyrrhizin over a surface with immobilized HMGB1.

Computational docking

Computational docking simulation was performed using RosettaDock 2.3 as previously described (Ref to JMB and PlosOne). The publically available experimental structures of CXCL12 (PDBid: 2kee) and HMGB1 (PDBid: 2yrq) were used for docking. The starting structures were visually oriented and then separated by 25 Å.

The CXCL12 was docked separately with BoxA and BoxB. The HMGB1 domains were kept fixed while CXCL12 was brought towards them with 3-8-8 perturbations [perturbations along the line of centers (in angstroms) – perturbation in the plane perpendicular to the line of center (in angstroms) – rotational perturbation (in degrees)].

After each docking run we obtained a total of 20,000 decoys. The lowest scoring models were visually analyzed and the decoys in accordance with the NMR mapping were selected for further refinement.

References

1 Marasco, W. A. & Sui, J. The growth and potential of human antiviral monoclonal antibody therapeutics. *Nat Biotechnol* 25, 1421-1434, doi:nbt1363 [pii]10.1038/nbt1363 (2007).

2 Sandercock, C. G. & Storz, U. Antibody specification beyond the target: claiming a later-generation therapeutic antibody by its target epitope. *Nat Biotechnol* 30, 615-618, doi:10.1038/nbt.2291nbt.2291 [pii] (2012).

3 Corti, D. & Lanzavecchia, A. Broadly neutralizing antiviral antibodies. *Annu Rev Immunol* 31, 705-742, doi:10.1146/annurev-immunol-032712-095916 (2013).

4 Zuiderweg, E. R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41, 1-7 (2002).

5 Kay, L. E. NMR studies of protein structure and dynamics. *J Magn Reson* 173, 193-207, doi:S1090-7807(04)00385-4 [pii]10.1016/j.jmr.2004.11.021 (2005).

6 Kay, L. E. NMR studies of protein structure and dynamics. 2005. *J Magn Reson* 213, 477-491, doi:10.1016/j.jmr.2011.09.009S1090-7807(11)00321-1 [pii] (2011).

7 Skrisovska, L. & Allain, F. H. Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. *J Mol Biol* 375, 151-164, doi:S0022-2836(07)01206-5 [pii]10.1016/j.jmb.2007.09.030 (2008).

8 Skrisovska, L., Schubert, M. & Allain, F. H. Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *J Biomol NMR* 46, 51-65, doi:10.1007/s10858-009-9362-7 (2010).

9 Michel, E., Skrisovska, L., Wuthrich, K. & Allain, F. H. Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology. *Chembiochem* 14, 457-466, doi:10.1002/cbic.201200732 (2013).

10 Clore, G. M. & Gronenborn, A. M. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252, 1390-1399 (1991).

11 Clore, G. M. & Gronenborn, A. M. Structures of larger proteins, protein-ligand and protein-DNA complexes by multi-dimensional heteronuclear NMR. *Prog Biophys Mol Biol* 62, 153-184 (1994).

12 Wagner, G. Prospects for NMR of large proteins. *J Biomol NMR* 3, 375-385 (1993).

13 Wider, G. & Wuthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr Opin Struct Biol* 9, 594-601, doi:S0959-440X(99)00011-1 [pii] (1999).

14 Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29, 4659-4667 (1990).

15 Frueh, D. P. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog Nucl Magn Reson Spectrosc* 78C, 47-75, doi:S0079-6565(13)00101-5 [pii]10.1016/j.pnmrs.2013.12.001 (2014).

16 Leopold, M. F., Urbauer, J. L. & Wand, A. J. Resonance assignment strategies for the analysis of NMR spectra of proteins. *Mol Biotechnol* 2, 61-93, doi:10.1007/BF02789290 (1994).

17 Jung, Y.-S. & Zweckstetter, M. Mars -robust automatic backbone assignment of proteins..pdf>. *Journal of Biomolecular NMR* 30, 11-23 (2004).

18 Jang, R., Gao, X. & Li, M. Towards fully automated structure-based NMR resonance assignment of (1)(5)N-labeled proteins from automatically picked peaks. *J Comput Biol* 18, 347-363, doi:10.1089/cmb.2010.0251 (2011).

19 Williamson, R. A., Carr, M. D., Frenkiel, T. A., Feeney, J. & Freedman, R. B. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry* 36, 13882-13889, doi:10.1021/bi9712091 (1997).

20 Ulrich, E. L. *et al.* BioMagResBank. *Nucleic Acids Res* 36, D402-408, doi:10.1093/nar/gkm957 (2008).

21 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* 73, 1-16 (2013).

22 Blech, M. *et al.* Molecular structure of human GM-CSF in complex with a disease-associated anti-human GM-CSF autoantibody and its potential biological implications. *Biochem J* 447, 205-215, doi:10.1042/BJ20120884 (2012).

23 Blech, M. *et al.* One target-two different binding modes: structural insights into gevokizumab and canakinumab interactions to interleukin-1beta. *J Mol Biol* 425, 94-111, doi:10.1016/j.jmb.2012.09.021 (2013).

24 Varani, L. *et al.* Solution mapping of T cell receptor docking footprints on peptide-MHC. *Proc Natl Acad Sci U S A* 104, 13080-13085, doi:0703702104 [pii]10.1073/pnas.0703702104 (2007).

25 Bain, A. D. Chemical exchange in NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* 43, 63-103, doi:10.1016/j.pnmrs.2003.08.001 (2003).

26 Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat Struct Biol* 7, 220-223, doi:10.1038/73331 (2000).

27 Mage, M. G. Preparation of Fab fragments from IgGs of different animal species. *Methods Enzymol* 70, 142-150 (1980).

28 Bird, R. E. *et al.* Single-chain antigen-binding proteins. *Science* 242, 423-426 (1988).

29 Walker, L. M., Bowley, D. R. & Burton, D. R. Efficient recovery of high-affinity antibodies from a single-chain Fab yeast display library. *J Mol Biol* 389, 365-375, doi:10.1016/j.jmb.2009.04.019 (2009).

30 Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125, 1731-1737, doi:10.1021/ja026939x (2003).

31 S.J. de Vries, S. J. *et al.* HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 69, 726-733, doi:10.1002/prot.21723 (2007).

32 S.J. de Vries, S. J., A.D.J. van Dijk, M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* 5, 883-897, doi:nprot.2010.32 [pii]10.1038/nprot.2010.32 (2010).

33 Ryabov, Y., Clore, G. M. & Schwieters, C. D. Direct use of 15N relaxation rates as experimental restraints on molecular shape and orientation for docking of protein-protein complexes. *J Am Chem Soc* 132, 5987-5989, doi:10.1021/ja101842n (2010).

34 Pedotti, M., Simonelli, L., Livoti, E. & Varani, L. Computational docking of antibody-antigen complexes, opportunities and pitfalls illustrated by influenza hemagglutinin. *Int J Mol Sci* 12, 226-251, doi:10.3390/ijms12010226 (2011).

35 Simonelli, L. *et al.* Rapid structural characterization of human antibody-antigen complexes through experimentally validated computational docking. *J Mol Biol* 396, 1491-1507, doi:10.1016/j.jmb.2009.12.053 (2010).

36 Simonelli, L. *et al.* Rational engineering of a human anti-dengue antibody through experimentally validated computational docking. *PLoS One* 8, e55561, doi:10.1371/journal.pone.0055561 (2013).

37 Tugarinov, V. *et al.* NMR structure of an anti-gp120 antibody complex with a V3 peptide reveals a surface important for co-receptor binding. *Structure* 8, 385-395 (2000).

38 Mueller, G. A., Smith, A. M., Chapman, M. D., Rule, G. S. & Benjamin, D. C. Hydrogen exchange nuclear magnetic resonance spectroscopy mapping of antibody epitopes on the house dust mite allergen Der p 2. *J Biol Chem* 276, 9359-9365, doi:10.1074/jbc.M010812200 (2001).

39 Grinstead, J. S., Schuman, J. T. & Campbell, A. P. Epitope mapping of antigenic MUC1 peptides to breast cancer antibody fragment B27.29: a heteronuclear NMR study. *Biochemistry* 42, 14293-14305, doi:10.1021/bi0301237 (2003).

40 Morgan, W. D., Lock, M. J., Frenkiel, T. A., Grainger, M. & Holder, A. A. Malaria parasite-inhibitory antibody epitopes on Plasmodium falciparum merozoite surface protein-1(19) mapped by TROSY NMR. *Molecular and biochemical parasitology* 138, 29-36, doi:10.1016/j.molbiopara.2004.06.014 (2004).

41 Wilkinson, I. C. *et al.* High Resolution NMR-based Model for the Structure of a scFv-IL-1 Complex: POTENTIAL FOR NMR AS A KEY TOOL IN THERAPEUTIC ANTIBODY DESIGN AND DEVELOPMENT. *Journal of Biological Chemistry* 284, 31928-31935, doi:10.1074/jbc.M109.025304 (2009).

42 Razzera, G. *et al.* Mapping the interactions between a major pollen allergen and human IgE antibodies. *Structure* 18, 1011-1021, doi:10.1016/j.str.2010.05.012 (2010).

43 Jazayeri, A., Dias, J. M. & Marshall, F. H. From G Protein-coupled Receptor Structure Resolution to Rational Drug Design. *J Biol Chem* 290, 19489-19495, doi:10.1074/jbc.R115.668251 (2015).

44 Singh, R., Singh, S. & Pandey, P. N. In-silico analysis of Sirt2 from Schistosoma mansoni: structures, conformations, and interactions with inhibitors. *J Biomol Struct Dyn* 34, 1042-1051, doi:10.1080/07391102.2015.1065205 (2016).

45 Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W., Jr. Computational methods in drug discovery. *Pharmacological reviews* 66, 334-395, doi:10.1124/pr.112.007336 (2014).

46 Alvarez Dorta, D. *et al.* The Antiadhesive Strategy in Crohn's Disease: Orally Active Mannosides to Decolonize Pathogenic Escherichia coli from the Gut. *Chembiochem* 17, 936-952, doi:10.1002/cbic.201600018 (2016).

47 Janin, J. *et al.* CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* 52, 2-9, doi:10.1002/prot.10381 (2003).

48 Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409-443, doi:10.1002/prot.10115 (2002).

49 Rini, J. M., Schulze-Gahmen, U. & Wilson, I. A. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* 255, 959-965 (1992).

50 Press, E. M., Givol, D., Piggot, P. J., Porter, R. R. & Wilkinson, J. M. The chemical structure of the heavy chains of rabbit and human immunoglobulin G (IgG). *Proc R Soc Lond B Biol Sci* 166, 150-158 (1966).

51 Wikstrom, M., Drakenberg, T., Forsen, S., Sjobring, U. & Bjorck, L. Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. *Biochemistry* 33, 14011-14017 (1994).

52 Putnam, F. W., Liu, Y. S. & Low, T. L. Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the alpha 1 heavy chain. *J Biol Chem* 254, 2865-2874 (1979).

53 Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 273, 927-948, doi:10.1006/jmbi.1997.1354 (1997).

54 Chothia, C. & Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196, 901-917 (1987).

55 Tramontano, A., Chothia, C. & Lesk, A. M. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol* 215, 175-182, doi:10.1016/S0022-2836(05)80102-0 (1990).

56 Marcatili, P., Rosi, A. & Tramontano, A. PIGS: automatic prediction of antibody structures. *Bioinformatics* 24, 1953-1954, doi:btn341 [pii]10.1093/bioinformatics/btn341 (2008).

57 Whitelegg, N. R. & Rees, A. R. WAM: an improved algorithm for modeling antibodies on the WEB. *Protein Eng* 13, 819-824 (2000).

58 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* 215, 403-410, doi:10.1006/jmbi.1990.9999S0022283680799990 [pii] (1990).

59 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680 (1994).

60 Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A. M. Antibody structure, prediction and redesign. *Biophys Chem* 68, 9-16 (1997).

61 Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A. M. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J Mol Biol* 275, 269-294, doi:10.1006/jmbi.1997.1442 (1998).

62 Shirai, H., Kidera, A. & Nakamura, H. Structural classification of CDR-H3 in antibodies. *FEBS Lett* 399, 1-8 (1996).

63 Ponder, J. W. & Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193, 775-791, doi:0022-2836(87)90358-5 [pii] (1987).

64 Sivasubramanian, A., Sircar, A., Chaudhury, S. & Gray, J. J. Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins* 74, 497-514, doi:10.1002/prot.22309 (2009).

65 Sircar, A., Kim, E. T. & Gray, J. J. RosettaAntibody: antibody variable region homology modeling server. *Nucleic Acids Research* 37, W474-W479, doi:10.1093/nar/gkp387 (2009).

66 Canutescu, A. A. & Dunbrack, R. L., Jr. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci* 12, 963-972, doi:10.1110/ps.0242703 (2003).

67 Wang, C., Schueler-Furman, O. & Baker, D. Improved side-chain modeling for protein-protein docking. *Protein Sci* 14, 1328-1339, doi:ps.041222905 [pii]10.1110/ps.041222905 (2005).

68 Kortemme, T., Morozov, A. V. & Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* 326, 1239-1259, doi:S0022283603000214 [pii] (2003).

69 Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* 35, 133-152, doi:10.1002/(SICI)1097-0134(19990501)35:2<133::AID-PROT1>3.0.CO;2-N [pii] (1999).

70 Dunbrack, R. L., Jr. & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 6, 1661-1681, doi:10.1002/pro.5560060807 (1997).

71 Warshel, A., Russell, S. T. & Churg, A. K. Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability. *Proc Natl Acad Sci U S A* 81, 4785-4789 (1984).

72 Camacho, C. J., Gatchell, D. W., Kimura, S. R. & Vajda, S. Scoring docked conformations generated by rigid-body protein-protein docking. *Proteins* 40, 525-537, doi:10.1002/1097-0134(20000815)40:3<525::AID-PROT190>3.0.CO;2-F [pii] (2000).

73 Cheng, T. M., Blundell, T. L. & Fernandez-Recio, J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins* 68, 503-515, doi:10.1002/prot.21419 (2007).

74 Schueler-Furman, O., Wang, C. & Baker, D. Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins* 60, 187-194, doi:10.1002/prot.20556 (2005).

75 Wang, C. *et al.* RosettaDock in CAPRI rounds 6-12. *Proteins* 69, 758-763, doi:10.1002/prot.21684 (2007).

76 Gray, J. J. *et al.* Protein-protein docking predictions for the CAPRI experiment. *Proteins* 52, 118-122, doi:10.1002/prot.10384 (2003).

77 Kwong, P. D. & Wilson, I. A. HIV-1 and influenza antibodies: seeing antigens in new ways. *Nat Immunol* 10, 573-578, doi:10.1038/ni.1746 (2009).

78 Sircar, A. & Gray, J. J. SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLoS Comput Biol* 6, e1000644, doi:10.1371/journal.pcbi.1000644 (2010).

79 Zlotnik, A. & Yoshie, O. Chemokines: a new classification system and their role in immunity. *Immunity* 12, 121-127 (2000).

80 Clore, G. M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A. M. Three-dimensional structure of interleukin 8 in solution. *Biochemistry* 29, 1689-1696 (1990).

81 Lodi, P. J. *et al.* High-resolution solution structure of the beta chemokine hMIP-1 beta by multidimensional NMR. *Science* 263, 1762-1767 (1994).

82 Skelton, N. J., Aspiras, F., Ogez, J. & Schall, T. J. Proton NMR Assignments and Solution Conformation of RANTES, a Chemokine of the C-C Type. *Biochemistry* 34, 5329-5342, doi:10.1021/bi00016a004 (1995).

83 Handel, T. M. & Domaille, P. J. Heteronuclear (1H, 13C, 15N) NMR assignments and solution structure of the monocyte chemoattractant protein-1 (MCP-1) dimer. *Biochemistry* 35, 6569-6584, doi:10.1021/bi9602270 (1996).

84 Crump, M. P., Rajarathnam, K., Kim, K. S., Clark-Lewis, I. & Sykes, B. D. Solution structure of eotaxin, a chemokine that selectively recruits eosinophils in allergic inflammation. *J Biol Chem* 273, 22471-22479 (1998).

85 Mizoue, L. S., Bazan, J. F., Johnson, E. C. & Handel, T. M. Solution Structure and Dynamics of the CX3C Chemokine Domain of Fractalkine and Its Interaction with an N-Terminal Fragment of CX3CR1†,‡. *Biochemistry* 38, 1402-1414, doi:10.1021/bi9820614 (1999).

86 Swaminathan, G. J. *et al.* Crystal Structures of Oligomeric Forms of the IP-10/CXCL10 Chemokine. *Structure* 11, 521-532, doi:10.1016/s0969-2126(03)00070-4 (2003).

87 Liwang, A. C., Wang, Z. X., Sun, Y., Peiper, S. C. & Liwang, P. J. The solution structure of the anti-HIV chemokine vMIP-II. *Protein Sci* 8, 2270-2280, doi:10.1110/ps.8.11.2270 (1999).

88 Allen, S. J., Crown, S. E. & Handel, T. M. Chemokine:Receptor Structure, Interactions, and Antagonism. *Annual Review of Immunology* 25, 787-820, doi:10.1146/annurev.immunol.24.021605.090529 (2007).

89 Gayle, R. B., 3rd *et al.* Importance of the amino terminus of the interleukin-8 receptor in ligand interactions. *J Biol Chem* 268, 7283-7289 (1993).

90 Moser, B., Wolf, M., Walz, A. & Loetscher, P. Chemokines: multiple levels of leukocyte migration control. *Trends Immunol* 25, 75-84, doi:10.1016/j.it.2003.12.005 (2004).

91 Veldkamp, C. T. *et al.* Monomeric structure of the cardioprotective chemokine SDF-1/CXCL12. *Protein Science* 18, 1359-1369, doi:10.1002/pro.167 (2009).

92 Veldkamp, C. T. The monomer-dimer equilibrium of stromal cell-derived factor-1 (CXCL 12) is altered by pH, phosphate, sulfate, and heparin. *Protein Science* 14, 1071-1081, doi:10.1110/ps.041219505 (2005).

93 Doranz, B. J. *et al.* Identification of CXCR4 domains that support coreceptor and chemokine receptor functions. *J Virol* 73, 2752-2761 (1999).

94 Veldkamp, C. T., Seibert, C., Peterson, F. C., Sakmar, T. P. & Volkman, B. F. Recognition of a CXCR4 Sulfotyrosine by the Chemokine Stromal Cell-derived Factor-1α (SDF-1α/CXCL12). *Journal of Molecular Biology* 359, 1400-1409, doi:10.1016/j.jmb.2006.04.052 (2006).

95 Farzan, M. *et al.* The role of post-translational modifications of the CXCR4 amino terminus in stromal-derived factor 1 alpha association and HIV-1 entry. *J Biol Chem* 277, 29484-29489, doi:10.1074/jbc.M203361200 (2002).

96 Martin, C. *et al.* Chemokines acting via CXCR2 and CXCR4 control the release of neutrophils from the bone marrow and their return following senescence. *Immunity* 19, 583-593 (2003).

97 Klein, R. S. & Rubin, J. B. Immune and nervous system CXCL12 and CXCR4: parallel roles in patterning and plasticity. *Trends Immunol* 25, 306-314, doi:10.1016/j.it.2004.04.002 (2004).

98 Dar, A. *et al.* Chemokine receptor CXCR4-dependent internalization and resecretion of functional chemokine SDF-1 by bone marrow endothelial and stromal cells. *Nat Immunol* 6, 1038-1046, doi:10.1038/ni1251 (2005).

99 Nagasawa, T. *et al.* Molecular cloning and characterization of a murine pre-B-cell growth-stimulating factor/stromal cell-derived factor 1 receptor, a murine homolog of the human immunodeficiency virus 1 entry coreceptor fusin. *Proc Natl Acad Sci U S A* 93, 14726-14729 (1996).

100 David, N. B. *et al.* Molecular basis of cell migration in the fish lateral line: role of the chemokine receptor CXCR4 and of its ligand, SDF1. *Proc Natl Acad Sci U S A* 99, 16297-16302, doi:10.1073/pnas.252339399 (2002).

101 Odemis, V. *et al.* Mice deficient in the chemokine receptor CXCR4 exhibit impaired limb innervation and myogenesis. *Molecular and cellular neurosciences* 30, 494-505, doi:10.1016/j.mcn.2005.07.019 (2005).

102 Wald, O. *et al.* Involvement of the CXCL12/CXCR4 pathway in the advanced liver disease that is associated with hepatitis C virus or hepatitis B virus. *Eur J Immunol* 34, 1164-1174, doi:10.1002/eji.200324441 (2004).

103 Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 272, 872-877 (1996).

104 Martin-Garcia, J., Cao, W., Varela-Rohena, A., Plassmeyer, M. L. & Gonzalez-Scarano, F. HIV-1 tropism for the central nervous system: Brain-derived envelope glycoproteins with lower CD4 dependence and reduced sensitivity to a fusion inhibitor. *Virology* 346, 169-179, doi:10.1016/j.virol.2005.10.031 (2006).

105 Vallat, A. V. *et al.* Localization of HIV-1 co-receptors CCR5 and CXCR4 in the brain of children with AIDS. *The American journal of pathology* 152, 167-178 (1998).

106 van der Meer, P., Ulrich, A. M., Gonzalez-Scarano, F. & Lavi, E. Immunohistochemical analysis of CCR2, CCR3, CCR5, and CXCR4 in the human brain: potential mechanisms for HIV dementia. *Experimental and molecular pathology* 69, 192-201, doi:10.1006/exmp.2000.2336 (2000).

107 Peng, H. *et al.* HIV-1-infected and/or immune activated macrophages regulate astrocyte SDF-1 production through IL-1beta. *Glia* 54, 619-629, doi:10.1002/glia.20409 (2006).

108 Amara, A. *et al.* HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication. *J Exp Med* 186, 139-146 (1997).

109 Tang, D., Kang, R., Zeh, H. J., 3rd & Lotze, M. T. High-mobility group box 1 and cancer. *Biochim Biophys Acta* 1799, 131-140, doi:10.1016/j.bbagrm.2009.11.014 (2010).

110 Stott, K., Watson, M., Howe, F. S., Grossmann, J. G. & Thomas, J. O. Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains. *J Mol Biol* 403, 706-722, doi:10.1016/j.jmb.2010.07.045 (2010).

111 Knapp, S. *et al.* The long acidic tail of high mobility group box 1 (HMGB1) protein forms an extended and flexible structure that interacts with specific residues within and between the HMG boxes. *Biochemistry* 43, 11992-11997, doi:10.1021/bi049364k (2004).

112 Hardman, C. H. *et al.* Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* 34, 16596-16607 (1995).

113 Stros, M., Ozaki, T., Bacikova, A., Kageyama, H. & Nakagawara, A. HMGB1 and HMGB2 cell-specifically down-regulate the p53- and p73-dependent sequence-specific transactivation from the human Bax gene promoter. *J Biol Chem* 277, 7157-7164, doi:10.1074/jbc.M110233200 (2002).

114 Stros, M., Muselikova-Polanska, E., Pospisilova, S. & Strauss, F. High-affinity binding of tumor-suppressor protein p53 and HMGB1 to hemicatenated DNA loops. *Biochemistry* 43, 7215-7225, doi:10.1021/bi049928k (2004).

115 Strichman-Almashanu, L. Z., Bustin, M. & Landsman, D. Retroposed copies of the HMG genes: a window to genome dynamics. *Genome research* 13, 800-812, doi:10.1101/gr.893803 (2003).

116 Calogero, S. *et al.* The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nat Genet* 22, 276-280, doi:10.1038/10338 (1999).

117 Taniguchi, N. *et al.* Stage-specific secretion of HMGB1 in cartilage regulates endochondral ossification. *Mol Cell Biol* 27, 5650-5663, doi:10.1128/MCB.00130-07 (2007).

118 Bustin, M. & Neihart, N. K. Antibodies against chromosomal HMG proteins stain the cytoplasm of mammalian cells. *Cell* 16, 181-189 (1979).

119 Bianchi, M. E. & Beltrame, M. Upwardly mobile proteins. Workshop: the role of HMG proteins in chromatin structure, gene expression and neoplasia. *EMBO Rep* 1, 109-114, doi:10.1093/embo-reports/kvd030 (2000).

120 Bonaldi, T. *et al.* Monocytic cells hyperacetylate chromatin protein HMGB1 to redirect it towards secretion. *EMBO J* 22, 5551-5560, doi:10.1093/emboj/cdg516 (2003).

121 Ito, I., Fukazawa, J. & Yoshida, M. Post-translational methylation of high mobility group box 1 (HMGB1) causes its cytoplasmic localization in neutrophils. *J Biol Chem* 282, 16336-16344, doi:10.1074/jbc.M608467200 (2007).

122 Lamkanfi, M. *et al.* Inflammasome-dependent release of the alarmin HMGB1 in endotoxemia. *J Immunol* 185, 4385-4392, doi:10.4049/jimmunol.1000803 (2010).

123 Scaffidi, P., Misteli, T. & Bianchi, M. E. Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* 418, 191-195, doi:10.1038/nature00858 (2002).

124 Hatada, T. *et al.* Plasma concentrations and importance of High Mobility Group Box protein in the prognosis of organ failure in patients with disseminated intravascular coagulation. *Thromb Haemost* 94, 975-979, doi:10.1160/TH05-05-0316 (2005).

125 Bianchi, M. E. & Manfredi, A. A. High-mobility group box 1 (HMGB1) protein at the crossroads between innate and adaptive immunity. *Immunol Rev* 220, 35-46, doi:10.1111/j.1600-065X.2007.00574.x (2007).

126 Mollica, L. *et al.* Glycyrrhizin binds to high-mobility group box 1 protein and inhibits its cytokine activities. *Chem Biol* 14, 431-441, doi:10.1016/j.chembiol.2007.03.007 (2007).

127 Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *J Mol Biol* 373, 503-519, doi:S0022-2836(07)01003-0 [pii]10.1016/j.jmb.2007.07.050 (2007).

128 Schiraldi, M. *et al.* HMGB1 promotes recruitment of inflammatory cells to damaged tissues by forming a complex with CXCL12 and signaling via CXCR4. *J Exp Med* 209, 551-563, doi:10.1084/jem.20111739 (2012).

129 Veldkamp, C. T. *et al.* Structural Basis of CXCR4 Sulfotyrosine Recognition by the Chemokine SDF-1/CXCL12. *Science Signaling* 1, ra4-ra4, doi:10.1126/scisignal.1160755 (2008).

130 Kofuku, Y. *et al.* Structural Basis of the Interaction between Chemokine Stromal Cell-derived Factor-1/CXCL12 and Its G-protein-coupled Receptor CXCR4. *Journal of Biological Chemistry* 284, 35240-35250, doi:10.1074/jbc.M109.024851 (2009).

131 Mukhopadhyay, S., Kuhn, R. J. & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol* 3, 13-22, doi:10.1038/nrmicro1067 (2005).

132 Nakao, S., Lai, C. J. & Young, N. S. Dengue virus, a flavivirus, propagates in human bone marrow progenitors and hematopoietic cell lines. *Blood* 74, 1235-1240 (1989).

133 Halstead, S. B. Pathogenesis of dengue: challenges to molecular biology. *Science* 239, 476-481 (1988).

134 Holmes, E. & Twiddy, S. The origin, emergence and evolutionary genetics of dengue virus. *Infection, Genetics and Evolution* 3, 19-28, doi:10.1016/s1567-1348(03)00004-2 (2003).

135 Sabin, A. B. Research on dengue during World War II. *Am J Trop Med Hyg* 1, 30-50 (1952).

136 Diaz, A. *et al.* Description of the clinical picture of dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS) in adults. *Bull Pan Am Health Organ* 22, 133-144 (1988).

137 Noisakran, S. & Perng, G. C. Alternate hypothesis on the pathogenesis of dengue hemorrhagic fever (DHF)/dengue shock syndrome (DSS) in dengue virus infection. *Exp Biol Med (Maywood)* 233, 401-408, doi:10.3181/0707-MR-198 (2008).

138 Kyle, J. L. & Harris, E. Global spread and persistence of dengue. *Annu Rev Microbiol* 62, 71-92, doi:10.1146/annurev.micro.62.081307.163005 (2008).

139 Rothman, A. L. Dengue: defining protective versus pathologic immunity. *J Clin Invest* 113, 946-951, doi:10.1172/JCI21512 (2004).

140 Kuhn, R. J. *et al.* Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* 108, 717-725 (2002).

141 Bressanelli, S. *et al.* Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J* 23, 728-738, doi:10.1038/sj.emboj.76000647600064 [pii] (2004).

142 Lin, C. W. & Wu, S. C. A Functional Epitope Determinant on Domain III of the Japanese Encephalitis Virus Envelope Protein Interacted with Neutralizing-Antibody Combining Sites. *Journal of Virology* 77, 2600-2606, doi:10.1128/jvi.77.4.2600-2606.2003 (2003).

143 Lok, S. M. *et al.* Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. *Nat Struct Mol Biol* 15, 312-317, doi:10.1038/nsmb.1382 (2008).

144 Khan, A. M. *et al.* Conservation and variability of dengue virus proteins: implications for vaccine design. *PLoS Negl Trop Dis* 2, e272, doi:10.1371/journal.pntd.0000272 (2008).

145 Beltramello, M. *et al.* The human immune response to Dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity. *Cell*

Host Microbe 8, 271-283, doi:S1931-3128(10)00279-9 [pii]10.1016/j.chom.2010.08.007 (2010).

146 Crill, W. D. & Roehrig, J. T. Monoclonal Antibodies That Bind to Domain III of Dengue Virus E Glycoprotein Are the Most Efficient Blockers of Virus Adsorption to Vero Cells. *Journal of Virology* 75, 7769-7773, doi:10.1128/jvi.75.16.7769-7773.2001 (2001).

147 Pierson, T. C., Fremont, D. H., Kuhn, R. J. & Diamond, M. S. Structural insights into the mechanisms of antibody-mediated neutralization of flavivirus infection: implications for vaccine development. *Cell Host Microbe* 4, 229-238, doi:10.1016/j.chom.2008.08.004 (2008).

148 Gray, J. J. *et al.* Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. *Journal of Molecular Biology* 331, 281-299, doi:10.1016/s0022-2836(03)00670-3 (2003).

149 Halstead, S. B. Neutralization and antibody-dependent enhancement of dengue viruses. *Adv Virus Res* 60, 421-467 (2003).

150 de Vries SJ, van Dijk M, Bonvin AM. 2010. The HADDOCK web server for data-driven biomolecular docking.*Nat. Protoc*.5(5): 883–897

151 de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AM. 2007. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 69(4): 726–733

152 Ryabov Y, Clore GM, Schwieters CD. 2010. Direct use of 15N relaxationrates as experimental restraints on molecular shape and orientation for docking of protein–protein complexes. *J. Am. Chem. Soc*.132(17): 5987–5989

153 Wilkinson IC, Hall CJ, Veverka V, Shi JY, Muskett FW, Stephens PE, Taylor RJ, Henry AJ, Carr MD. 2009. High resolution NMR-based model for the structure of a scFv-IL-1 complex: potential for NMR as a key tool in therapeutic antibody design and development. *J Biol Chem* 284 (46): 31928–31935