Manuscript

Click here to download Manuscript
REVIEW_TRENDS.v6.FINALISSIMO-Trends.docx

Structure and computation in immunoreagent design: from diagnostics to vaccines

Louise Gourlay[1], Claudio Peri[2], Martino Bolognesi[1,3,*] and Giorgio Colombo[2,*]

[1]Dipartimento di Bioscienze, Università di Milano, 20133, Milan, Italy

[2]Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, Via Mario Bianco, 9, 20131, Milan, Italy.

[3]Centro di Ricerca Pediatrica Romeo ed Enrica Invernizzi, Università di Milano, Milan, Italy

* Authors to whom correspondence should be addressed:

Dr. Giorgio Colombo: g.colombo@icrm.cnr.it

Prof. Martino Bolognesi: martino.bolognesi@unimi.it

Abstract

Novel immunological tools for efficient diagnosis and treatment of emerging infections are urgently required. Advances in the diagnostic and vaccine development fields are continuously progressing, with Reverse Vaccinology and Structural Vaccinology methods for antigen identification and structure-based antigen (re)design playing increasingly relevant roles. Structural vaccinology, in particular, is predicted to be the front-runner in the future development of diagnostics and vaccines targeting challenging diseases such as AIDS and cancer.

We review state-of-the-art methodologies for structure-based epitope identification and antigen design, with specific applicative examples. We highlight the implications of such methods for the engineering of biomolecules with improved immunological properties, potential diagnostic and/or therapeutic uses, and discuss the perspectives of structure-based rational design for the production of advanced immunoreagents.

Keywords:

Structural vaccinology, *in silico* epitope discovery, epitope-driven vaccine design, antigen engineering, immunoreagents

Introduction

Together with improved nutrition and hygiene, efficient disease diagnosis and vaccination are medicine landmarks that have promoted major improvements in human health and longevity [1, 2]. In diagnostics, significant progress has been observed in the development of disease-specific probes, in the types of materials used for their presentation, and in the sensitivity of the associated detection methods. It is now possible to design new (bio)molecules that recognize specific **biomarkers** (see Glossary), such as **antibodies** (Abs) elicited in response to infection, and improve our understanding of the molecular mechanisms underlying the immune response [3]. The impacts of such advances are particularly visible in fields where diagnosis typically required invasive techniques or complex scanning machineries. In parallel, vaccines offer great potential to tackle infectious diseases caused by emerging pathogens, multi-drug resistant bacteria, parasites and viruses, and lately in the treatment of cancer and HPV, or HIV and even Ebola [4-7]. Vaccines help prevent and/or cure diseases by activating the immune system against (future) attacks. A fundamental step in the delivery of protective vaccines has been, and is, the development of new approaches to the discovery and engineering of **protein antigens** (Ags) capable of eliciting apt immune responses.

Epitopes, the specific molecular regions of an antigen that are recognized in the immune response, may represent ideal immunoreagents that can be conjugated to other known immunogens (proteins and otherwise), or presented in combination with other epitopes on chemical or biological scaffolds, increasing their applicative potential as diagnostics and vaccine components [8]. A holistic approach, integrating both diagnostic probe and vaccine candidate design, is summarized in Figure 1.

Here, we review novel avenues and perspectives for Ag/epitope discovery and engineering in immunodiagnostics and vaccinology, based on the combination of protein 3D structures, *in silico* analyses and design strategies. We discuss specific applicative examples and highlight their implications for engineering synthetic molecules with improved immunological properties for diagnostic and/or therapeutic applications. Notably, although much of the research discussed originated from the need to combat specific infectious diseases, the scope of the approaches we present is vast, and in principle transferable to any disease-related Ag of known 3D structure.

Bioinformatics antigen selection: the Reverse Vaccinology revolution

In contrast to classical methods based on attenuated or killed pathogen particles to prompt the host immune response, modern vaccinology can rely on Ag discovery via bioinformatics; **Reverse Vaccinology** (RV; Box 1) explores pathogen genomes to identify Ags for downstream *in vivo* validation as vaccine candidates [9, 10]. Potential Ag genes are picked out from the bacterial genome *via* bioinformatic mining for specific sequence motifs, such as those encoding for outer membrane proteins, and then produced in an appropriate expression system (e.g. *Escherichia coli*) for testing. The extension of RV to explore multiple related pathogenic genomes, so-called Pan-genomic RV (PRV), can support the discovery of candidate Ags that cross-protect against diverse bacterial strains. The impact of RV has been exemplified by the identification of four protective Ags that formulate Bexsero^{R}, the first vaccine targeting *Neisseria meningitidis* serotype B infections [11].

Since bioinformatics analyses may yield large numbers of candidates, to reduce the number of proteins to be expressed-purified for *in vivo* phases, further methodological developments have been proposed, taking into account recurring structural/functional features (protective signatures) of known protective Ags, in order to better select bactericidal molecules [12]. When applied to *Staphylococcus aureus* and Group *B Streptococcus*, such 'protectome' analyses identified two previously unknown Ags, and 'rediscovered' known ones [12].

The role of structural studies in Ag engineering

Ag discovery can be efficiently supported by structure-based design concepts. In this context, **Structural Vaccinology (SV)** (see Glossary and Box 1), whereby 3D information is used to reveal the determinants of Ag stability and conformational dynamics underlying molecular recognition by the immune system, has recently emerged as a new driving force in immunology. Advances in X-ray crystallography, **Nuclear Magnetic Resonance** (NMR) and single **particle Cryo-Electron Microscopy** (Cryo-EM), are in fact facilitating the study of challenging Ags, particularly when applied to Ag and Ag-Ab complex structures and interaction studies [9, 13, 14].

Coupled to *in vitro* epitope mapping data and computer-aided Ag engineering, 3D structural knowledge can pinpoint the location and physico-chemical properties of the **epitopes** that are ultimately responsible for triggering B-cell and/or T-cell

mediated immune responses [15] (Figure 1). Structural knowledge on epitopes can also drive the design of peptide-based mimics (or libraries), epitope-containing protein domains, or completely new Ags that on one hand are reactive versus disease biomarkers, and on the other can elicit bacteria-neutralizing responses [9, 13]. Successful achievements of SV include applications to the Respiratory Syncytial Virus fusion protein Ag and the Factor H binding protein (FHbp) Ag from MenB, respectively [11, 16].

To progress along such promising avenues, we must address fundamental questions: 1) can we formulate general principles that allow immune-reactive and non-reactive epitope substructures to be distinguished?; 2) is it possible to design peptide-based epitope mimics (to be presented in multiplex or multivalent formats) as diagnostic probes that detect specific biomarkers?; 3) is it possible to translate epitope knowledge into the design of 'super-Ags' presenting multiple T-/B-cell epitopes that induce better protection compared to their single epitope counterparts? In fact, epitope-based vaccines can offer several advantages: they may be more readily produced on an industrial scale, are easily presented through diverse delivery vessels and **adjuvants**, and may be engineered to improve their antigenicity. In diagnostic applications, the possibility of displaying multiple reactive epitopes from one pathogen can facilitate the transition towards high-throughput approaches.

Epitope discovery and design

The first prerequisite to meet the above challenges is our ability to predict and locate epitope regions as privileged Ag-substructures capable of selectively engaging molecules of the immune system.

a) Epitope predictors based on general molecular properties

Computational methods for epitope prediction are playing increasingly important roles, thanks to continuous technological advances in computing. Moreover, a number of databases such as the Immune Epitope DataBase (www.iedb.org) [17], AntiJen [18], and the international ImMunoGeneTics (IMGT) [19] have been devised to collect experimental epitope sequence data, and to provide tools for the discovery of new epitopes, whilst providing information on those already mapped.

Although the precise criteria that determine immunogenic properties remain elusive, a number of molecular features shared by effective epitopes have been observed. Examples include localized physico-chemical properties, such as hydrophobicity, solvent-exposure, flexibility or lack of sequence conservation. Such observations led to the development of several computational algorithms. Notable examples of servers that predict B-cell epitopes through the combination of different molecular properties into a single predictor include ElliPro [20] (http://tools.immuneepitope.org/ellipro/) and SEPPA[21] (http://lifecenter.sgst.cn/seppa/).

b) Sequence-based epitope predictors

Due to the linear nature of T-cell epitopes, *in silico* sequence-based methods that predict both MHC class I and, to a lesser-extent, MHC class II molecule presentation, can be highly accurate. Recent examples include MULTIPRED [22], TEPITOPE (Human Leukocyte Antigen (HLA) class I) [23], and ProPred [24] (HLA class II predictor). It must be underlined that these approaches identify epitopes based on their predicted HLA binding affinities, and no information on their potential immunogenicity is revealed.

Predictions of B-cell epitopes, targeted by Abs in the initial stages of the immune response, are more challenging, since 3D structure information is typically required. Sequence-based methods were previously limited to the prediction of linear epitopes, however tools for the prediction of *conformational* epitopes, which represent a large subset of all B-cell epitopes, including many capable of inducing bactericidal responses [25], have recently emerged thanks to machine learning approaches [26, 27]

c) Structure-based epitope predictors

As the interplay between sequence and conformational properties underpins Ab binding, prediction methods based on explicit consideration of Ag structural dynamics have great potential for epitope discovery and design. One such method, called the Matrix of Low Coupling Energies (MLCE), implemented in a webserver (BEPPE, Binding Epitope Prediction from Protein Energetics, http://bioinf.uab.cat/BEPPE/)[28] analyzes the dynamic and energetic properties of Ag structures, and identifies contiguous surface-located residues that build low

intensity interaction networks with designated core protein regions [29]. Such substructures are not required for global stabilization of the Ag, may tolerate mutations and may be highly dynamic, all hallmarks of typical B-cell epitopes. MLCE, whose predictions reach promising accuracy (AUC 0.71), has been tested against Ags from diverse bacterial pathogens. Successful application to *Chlamydia* and *Burkholderia pseudomallei* Ags led to the identification of epitopes with both therapeutic and diagnostic implications [30-33]. MLCE/BEPPE can also be combined with energy domain decomposition methods to fragment proteins into sub-domains [34] prior to epitope predictions, revealing 'hidden epitopes' that may evade identification but be potentially interesting in generating immune responses [35] (Figure 2). Importantly, these energy-based strategies proved successful for identifying *conformational* epitopes. This task is now becoming increasingly feasible thanks to the soaring numbers of Ag and Ag-Ab 3D complex structures in the Protein Data Bank (http://www.rcsb.org).

The accuracy of the fore-mentioned methods may be further enhanced by combining them with approaches that focus on diverse physico-chemical properties. As for protein-protein interactions in general, Ab-Ag binding involves displacement of water molecules bound at the interaction interfaces: the Electrostatic Desolvation Profiles (EDP) method scans the protein surface, looking for solvent-accessible clefts and calculates the theoretical desolvation penalty for each site, upon binding of a protein partner, selecting those where water displacement is favored [36].

Synthetic epitope mimics: structure-activity relationships.

Information on the immunoreactive potential of epitopes coupled to the synthetic accessibility of peptides opens the way to chemical interventions aimed to modulate their structures and to assess how specific conformational ensembles affect Ab-recognition. One approach is to constrain peptides into the desired conformation by means of covalent modifications, such as stapling and macrocyclization [37-39]. Ideally, a stapling strategy may be a versatile tool to produce synthetic epitopes that closely mimic the conformational space of their native counterparts, such as loops and turns. For instance, the correct spacing between extremities of a loop can be achieved through a suitable macrocyclization process, or through the inclusion of PEG moieties mimicking the spatial distance between them [35].

This rationale was applied to a computationally-predicted epitope (Pal3) from the pathogenic bacterium *Burkholderia pseudomallei* (Figure 3) that was found to be strongly immunoreactive [40]. In the cognate Ag, the epitope adopts an α-helical conformation; when synthesized as a free peptide, in contrast to the full-length protein, the epitope mimic could clearly discriminate between melioidosis patient serogroups, and elicited bactericidal Abs. In the free peptide form, however, the epitope was unable to retain its α-helical conformation and, accordingly, exhibited poor stability in plasma. A 1,4-disubstituted-1,2,3-triazole staple was introduced to block the α-helical fold of Pal3, which improved both its half-life in human plasma and patient IgG recognition. In contrast, the Abs generated against the stapled peptide proved significantly less bactericidal than those elicited against the unconstrained epitope [40]. This observation was reconciled considering that, when housed in the recombinant Ag, the epitope region may undergo local folding-unfolding reactions that present different conformations to the polyclonal pools of generated Abs. Blocking a single conformation in the stapled synthetic epitope may thus give rise to a limited pool of Abs that recognize fewer epitope conformations. Such restrictions were not observed for the unstapled version of Pal3. In contrast, conformational preorganization of the synthetic epitope was found to improve its diagnostic performance, most likely by increasing affinity for specific Abs, translating into better readable signals. While the Pal3 case was one of the first reports of a stapling strategy applied to a peptide with diagnostic applications, these results suggest that the balance between conformational freedom and immunogenicity deserve more extended investigations [30, 40].

An additional notable example of epitope stapling involves the SAH-MPER linear epitope from the HIV gp41 protein. The constrained epitope was remarkably resistant to proteolysis, and was able to recapitulate its native antigenic properties by binding neutralizing Abs with high affinity [39]. The stapling technique used in this case involved the direct linkage of two carbon atoms within the epitope.

Multi-presentation of designed epitopes

Peptides mimicking Ab-recognition motifs of different Ags are ideal tools for multiplex approaches, whereby multiple epitopes can be presented simultaneously, both for detection of (multiple) biomarkers and for the induction of improved immune

responses (Figures 1,4; Box 2). The potential of peptide multiplex display methods has recently been demonstrated [41-43]. **Multiplex approaches** can significantly contribute to the serological detection of infections, and can discriminate between Ab classes (IgM, IgG, IgA), thus providing crucial information on infection type and disease progression (*i.e.* acute or chronic phases). The combination of responses originated from probe diversity/redundancy can facilitate the analysis and interpretation of results and ultimately can shed light on infection status, identify patient subgroups, discriminate between different pathogens (and their variants) and assist in medical decision-making. A further advantage is that multiplex diagnostics, based on large peptide libraries spotted on a single chip, are practically simple and therefore can be used on-site in hospitals, offering fast and accurate diagnosis that accelerates therapeutic decisions. This is extremely advantageous for low-income countries where infection outbreaks may be difficult to contain, but also in developed countries where new and re-emerging infections are on the increase, as underlined by recent pandemics (*e.g.* Ebola and Zika viruses).

In this framework, Kodadek *et al.* have recently demonstrated the use of **peptoid** libraries to mimic the molecular shapes in antigen space, to map the antigen-binding pockets of disease-specific Abs [3]. In one notable example, by screening around a million compounds, the authors found a subgroup that preferentially bound human IgGs from HIV positive patients. Reactive peptoids were re-synthesized and singularly evaluated: the most responsive one allowed serodiagnosis of HIV [44].

Another case reported the detection of *Burkholderia cepacia complex* infections from sera of Cystic Fibrosis (CF) patients using a peptide-based microarray platform [32]. In this case, peptide-mimics of predicted Ag-Ab binding interfaces were identified and chemically optimized for click-chemistry controlled orientation on the microarray surface. The resulting platform, displaying multiple *Burkholderia*-related epitopes was able to correctly diagnose infected individuals even in the face of superinfections caused by other prevalent CF pathogens [32]. The time and cost for the analysis were contained, and the results demonstrated that a peptide microarray approach may elucidate *Burkholderia sp.* infection even at the subtype level.

Multi-presentation: from nanoparticles to structure-based epitope grafting

One of the most exciting perspectives for computational epitope engineering is the development of new vaccines, and of novel epitope delivery vessels as agents

promoting the immune response. A possible multiplex approach based on nanocarriers is illustrated in Figure I, Box 3

These carriers can be mainly of two kinds: inorganic (*e.g.* gold) nanoparticles (AuNPs) functionalized with antigenic epitopes, and vesicular nanoparticles, composed of vesicles or liposomes, capable of carrying the Ags on their surface or inside the bilayer. Products based liposomal carriers are already commercial, such as Inflexal® V against influenza and Epaxal® against Hepatitis A [45]. One of the advantages of this technology is that lipid composition can be adjusted to meet desired properties, including size to accommodate the cargo, charge (charged liposomes are able to enhance the innate immune response stimulation [46], fluidity, or the inclusion of immunostimulatory lipids such as CAF01 [47] or α-galactosylceramide [48].

Concerning epitope presentation and delivery, liposomes may accommodate different epitopes encapsulated in the aqueous core or displayed at the surface. Epitope attachment can be mediated by covalent- or hydrophobic-anchors adsorbed through the lipid layer. Alternatively, the epitope may be designed and synthesized to include a transmembrane helix to promote liposome [49] attachment. Such versatility allows optimal display of synthetic epitopes: for instance, they may be presented directly to immature B-cells adsorbed or linked at the surface of the nanocarriers. On the other hand, T-cell epitopes may be encapsulated within the bilayer, and optimized in size and composition to be internalized by Ag-presenting cells, where the nanovesicles can release their cargo. It is known that different vesicles may promote uptake by different cellular types, so to favor MHC-I or MHC-II presentation [50-52]. One such approach was tested with good results by Irvine and coworkers, who simultaneously displayed the membrane proximal external region from the B-cell HIV epitope gp41 on the surface of the liposome, with a T-cell epitope (HIV-30) encapsulated within the internal region of the liposome [53]. Other advantages of using vesicle-based carriers include good tolerability and safety; importantly, they are completely biodegradable (Figure I, Box 3).

Evolving these technologies for vaccine purposes is a hot research topic [52]. Even if a protocol that gives rise to neutralizing protection has yet to be demonstrated, some encouraging results have been achieved [54]. Such studies hint at the possibility of using nanoparticles not only as viable carriers, but also to promote a boost in

immunogenicity compared to the use of the individual Ags, a common plus of nanocarriers. This aspect introduces further potential for rational engineering of the immunological response. Liposomes and nanoparticles can be in fact prepared to display a large diversity of epitopes, recapitulating the immunogenic determinants of many Ags into a single component. Also, the carriers can be optimized for specific tasks: diffusion, stability, stimulation of one specific branch of the adaptive immune response, act as cytokine carriers, act as adjuvants, or they can be designed to carry over many tasks at the same time. Compared to the typical subunit vaccine, the increase in complexity is high but the technological gain and the possibilities that they open are highly exciting.

Chimeras

Among the strategies that employ the rational design of Ags and epitopes, we consider the case of "chimeras" as a particularly interesting one [16] (Figure 4). One Ag can be characterized for sequence similarity across multiple pathogenic variants; then, assisted by 3D structural data, it is possible to assess the differences that define the variant-specific neutralizing epitopes. A series of mutations that encompass the regions of variability as well as specificity, can lead to a single recombinant Ag able to trigger a neutralizing effect in multiple strains, as successfully accomplished for factor H binding protein (fHbp) from *Neisseria meningitidis* group B (MenB) [55]. In this case, a broadly reactive Ag was requested in order to induce protection against all major MenB strains and, along with the other Ags in the formulation, to provide an effective vaccine. The objective was met by screening many different variants and mutations; however, this process may benefit from additional knowledge on the structural features of Ag-Ab recognition (*e.g.* Ags co-crystallized with FABs of selective Abs). In this regard, computational studies integrating sequence conservation across variants (bioinformatics) and Ag-Ab binding modeling and optimization (computational chemistry) or crystal structures, may help focus on a smaller set of "hotspot" residues within the epitope, thus reducing residue permutations to those that are more prevalent across the strains of interest, eventually minimizing the number of screened mutations.

Epitope Grafting

Computational design of proteins to host foreign functional motifs can give rise to Ags with new functionalities or diverse immunological properties. This strategy, termed ‘epitope grafting’, can be used to selectively stabilize suitable conformations of epitopes for immune presentation or to address antigenic variability of diverse pathogenic species (Figure 5, Box 4). Simultaneous presentation of multiple epitopes on the same scaffold is known to induce stronger and more durable immune response. For some pathogens, in particular of intracellular nature, joint cell-mediated and humoral immune responses are required for a sterilizing vaccine. It has been shown that immunization with chronic and acute phase *Mycobacterium tuberculosis* Ags on the same polypeptide chain, induces a better protective response in comparison to administering both Ags separately [56]. As an alternative to expressing the full-length forms of multi-Ags on the same polypeptide chain, epitope-grafting strategies could be adopted to combine both T-cell and B-cell epitopes within the same protein scaffold.

For successful epitope grafting, a main prerequisite is the availability of a protein scaffold that presents a 3D structure reflecting or compatible with that of the target Ab recognition motif to be transplanted. The field is evolving rapidly, but it has yet to reach its full potential. Using the Fold From Loops (FFL) method that designs scaffold proteins hosting new functional motifs, a neutralizing epitope from human respiratory syncytial virus (RSV) glycoprotein was grafted onto an unrelated protein scaffold, which elicited neutralizing Abs [57]. A recent report describing an automated tool for the implantation of immunogenic epitopes onto a given scaffold, named SAGE (Strategy for Alignment and Grafting of Epitopes), was published. The method considers both sequence and structure conservation between target epitopes and scaffolds, and is fully applicable to any protein from any pathogen for which the epitope location within the Ag 3D structure is known [58].

Concluding Remarks

Health challenges posed by globalization underline the growing request for delivery of protective vaccines at immediate notice; recent examples of viral outbreaks in Central Africa are just one of many examples. Antigens/epitopes that induce protective antibodies are the main components in vaccine formulations; nevertheless, classical methods of antigen selection cannot cope with such urgent

therapeutic requests. A new, more rational avenue involving the *in silico* screening of pathogen genomes, and taking advantage of our theoretical knowledge on protein structure, dynamics and recognition, provides a more rapid and economical alternative to conventional antigen discovery approaches. Further to antigen selection, *in silico* methods have also been demonstrated to accurately predict epitopes from sequence and 3D structure information, adding a powerful arm to vaccine design strategies. Despite such initial achievements, and a few successful application cases, the full potential of SV has yet to be exploited (see Outstanding Questions Box). In fact, general structural rules and the general applicability of antigen engineering approaches, such as chimeras or epitope grafting on existing protein architectures, are foreseen in the near future. Furthermore, antigen/epitope display on carriers such as nanoparticles, vesicles or liposomes are increasingly attractive approaches to vaccine design. The current technological control of the production of such decorated carrier species must be adequately met by systematic *in vivo* analyses of their immunologic and protective properties.

Health challenges must also be faced through adequate diagnostic capabilities. Due to extensive migration of world populations, pathogen diffusion between continents is increasing, calling for rapid, robust and practically facile diagnostic tests for use in hospitals and refugee camps/centers to diagnose, and contain, emerging infections. In this context, protein antigens/epitopes can be used to detect the presence and phase of pathogen infection by monitoring their recognition by cognate serum antibodies. The predictive capabilities acquired through the application of fundamental principles of protein structure recognition are therefore translating into new bio-reagents for the assembly of diagnostic tools based on microarray display technologies. Here, engineering of antigens/epitopes in a multiplex context is the most promising approach, given the possibility of multiple diagnostic queries on the same support, which can aptly translate in an efficient profiling of the disease and disease state, thus favoring patient stratification and therapeutic decision-making.

As in other branches of life sciences, genomic data and fundamental information on protein structure and dynamics are fostering the development of translational science. It is expected that vaccines and immunodiagnostics will soon rely on faster discovery and production methods, and on the broad distribution of safe-stable bio-reagents.

Acknowledgements

We acknowledge the following grant agencies for financial support: PRIN 2015 contract number 2015JTL4HL.

Glossary

Adjuvant: a substance that potentiates an immune response, and a fundamental component of vaccine formulations.

Antibody: An antibody (or immunoglobulin) is a protein composed of two heavy- and two light-chains that is secreted by B-cells in response to an antigen.

Antigen: A protein antigen is a pathogen-derived protein that triggers the host immune response. Antigens can also be based on oligosaccharides or other macromolecules.

Antigenic variability: the interspecies-dependent sequence variability of an epitope region from the same antigen, often acquired by a pathogen to evade the host immune response.

B-cell: Antibody-secreting cell, and key player in humoral immunity; also termed B-lymphocyte.

Bioinformatics: computational tools developed and applied to analyze data provided by genomic or protein sequencing projects.

Biomarker: a biological molecule that is characteristic of a given physiological or pathological process; it can be exploited for detection/diagnosis.

Cell-mediated immune response: the human immune response branch mediated by phagocytic cells, T-cells and cytokines.

Click-chemistry: the introduction of chemical staples to constrain the 3D conformation of a macromolecule.

Epitope: the antigen portion that stimulates an immune response. It stimulates T-cells or B-cells, depending on the nature of the epitope.

Epitope grafting: computer-guided transplantation of a (conformational) epitope onto a heterogeneous protein scaffold.

Half-life: the time required for an entity to decrease to half its initial value.

Human Leukocyte Antigen complex: HLA, a cell surface protein complex governing the human immune response and human equivalent to the major histocompatibility complex.

Humoral immune response: the branch of the human immune response triggered by antigens or pathogens circulating in the blood, which are recognized by B-cells and stimulate antibody secretion from plasma cells.

Major Histocompatibility Complex: MHC, Protein complex present on the cell-surface in all vertebrates that helps the immune system recognize foreign substances by binding to antigens and presenting them to T-cells.

Multiplex: the presentation of multiple biomarkers on a suitable vessel.

Neutralizing Abs: Antibodies that eradicate a pathogen from the body.

Nuclear Magnetic Resonance: Spectroscopy-based analytical technique used to determine 3D structures by exploiting the magnetic properties of atomic nuclei arranged within a (macro)molecule.

Peptoid: poly-N-substituted glycine residues that are protease-resistant peptide mimics.

Reverse Vaccinology: a bioinformatics-based approach to Ag identification, starting from pathogen genome sequence data.

Structural Vaccinology: the use of 3D structure information to guide specific engineering of antigens endowed with improved biochemical and/or immunological properties.

T-cell: Front-runner in cell-mediated immunity. There are several types of T-cell (or T-lymphocyte) that trigger and facilitate diverse types of immune response.

X-ray crystallography: A biophysical technique that reveals the 3D arrangement of atoms in a crystallized (macro)molecule, based on the application of X-rays whose wavelength approximately matches the typical dimensions of atomic level features, which are therefore resolved in the electron density maps.

Text Boxes

- **BOX 1: Reverse Vaccinology and Structural Vaccinology in Vaccine Development**

  Since the genomic advent in 2000, which led to increasing pathogen genome sequence information, bioinformatics-based methods for antigen selection (termed Reverse Vaccinology, RV) have taken over laborious and time consuming conventional approaches that require pathogen cultivation [10, 59]. RV methods predict surface-exposed or secreted proteins that are the first point of contact for immune system molecules as potential antigens, based on the presence of signal peptides and/or other sequence motifs that imply cell surface location, *e.g.* transmembrane regions, sortase recognition motifs, and sequence conservation

with known antigens. When multiple pathogen species genomes are available, RV can be extended to identify proteins pertaining to the pan-genome, shared by all species, whose common antigens could potentially protect against more than one species. *In silico* selected candidate antigens are subsequently validated for their effective immunological properties *in vitro* and in *in vivo* animal models.

Since protein antigen immunogenicity is directly determined by overall three-dimensional fold, conformations and related dynamics, unsurprisingly, structure-based antigen design followed closely on the heels of RV (originating Structural Vaccinology, SV). Based on 3D structure information, structural modifications such as the introduction of disulfide bonds can be strategically introduced to block typically transient conformations of neutralizing epitopes or antigens, as exemplified by the Respiratory Syncytial Virus (RSV) glycoprotein F antigen [15]. Alternatively, SV can be used to design epitope-containing domains that may be easily produced in recombinant form for industrial-scale production, as an alternative to the full-length antigen. Moreover, mutations can be introduced in epitope regions to account for interspecies sequence variability to generate cross-protective antigens. In short, SV can be used to specifically tailor protein antigens that can be rapidly produced for immediate testing of their biochemical and/or immunological properties.

- **BOX 2: Multiplex diagnostics based on microarray approaches.**

The distinctive trait of microarray technologies is the simultaneous detection of a panel of different molecules in a single experiment. In general, in analytical microarrays, target bait molecules are immobilized on a surface, creating a collection of probes that are able to "fish" complementary protein targets out of a biological sample [60, 61]. Unbound molecules are then removed by washing, and captured ligands are detected using labeled reporters, typically secondary antibodies. Arrays consist of a rigid support (e.g. a glass microscope slide) modified by suitable surface chemistry to immobilize bait probes in microspots of 100-250 μm diameter, spatially ordered in matrix format.

For antibody and biomarker detection, proteins and peptides are immobilized as capturing agents, whereas a labeled anti-immunoglobulin is used for detection. This is the scheme typically used for allergy tests (to detect allergen specific immunoglobulins E), infectious disease diagnosis (pathogen specific immunoglobulins G or M detection) and autoimmunity diagnostics. In this

context, peptide microarrays offer potentially unlimited flexibility as they can be designed to identify any antibody-binding domain that can subsequently be implemented in vaccine development. Design of epitope collections from RV- or SV-selected antigens may generate a large amount of molecular and immunological information by mining the Ab content present in plasma of individual patients, profiling immuno-responses to infection, vaccination etc. Importantly, peptides (relatively economical to produce and more managable compared to DNA or proteins) in a miniaturized and multiplexed format test can simultaneously screen for dozens to thousands of biomarkers in a single assay. Different chemistries for surface display are now available. Epitope-based microarrays exploit thus the structural and recognition properties of epitopes from target protein antigens to detect specific antibody-biomarkers. The main advantage in this kind of approach resides in the ability to comparatively analyze results from libraries of baits while exploiting the combination of responses originated from probe redundancy.

The maturity of the technique has been proven by the many published clinical applications and by the several commercial providers offering support and services.

- **BOX 3: Nanocarriers for epitopes**

Chemistry at the nanoscale is providing useful approaches to address biologically-relevant problems (Figure I). Particles and molecular machines of nm dimensions are being increasingly and more extensively used as drug-delivery tools for targeted therapeutics, for instance in cancer applications. In the context of vaccine delivery systems, nanotechnology can provide new solutions for the development of safe and effective products. In principle, by linking several different antigens (even of different chemical origin, e.g. proteins and carbohydrates) on the same carrier, it is possible to create highly reactive systems that may provide strong and durable protection: several studies have proven this potential, by encapsulating antigens to improve their stability or to promote presentation by displaying antigens on the nanoparticle surface. In this context, gold nanoparticles (AuNPs) have generated huge interest thanks to their biocompatibility, low toxicity, imaging properties and the ability to release antigens at specific sites through targeting [62]. In the case of peptide-based epitope mimics, the facile approaches to link peptides to the surfaces based on

sulfur-gold chemical reactivity allows for the versatile inclusion of structurally diverse sequences.

In the nanoparticle field, Outer Membrane Vesicles (OMVs) represent an innovative technology that allows several Ags (of *e.g.* bacterial or cancer origin) to be expressed and displayed on the vesicle surface. OMVs are heterogenous spheroid particles (10–300 nm in diameter) generated through the budding out of the bacterial outer membrane in Gram-Negative bacteria. OMVs purified from several pathogens proved to induce protective immune responses against the pathogens they derive from. Importantly, anti-*Neisseria* OMV-based vaccines are already available for human use [63-65]. Such remarkable protection can be linked to the fact that: they carry surface-associated protective antigens; they are readily phagocytosed by antigen-presenting cells, favoring presentation of antigen-derived peptides; and OMVs carry pathogen-associated-molecular patterns which target pathogen recognition receptors stimulating both innate adaptive immune responses. In the context of rational vaccine design, a key feature of OMVs is the possibility to manipulate their protein content and surface display by genetic engineering, directing protein expression to the outer membrane of the OMV-producing strain. This would clearly favor to development of highly effective multi-valent vaccines. Finally, such nanoparticles are gaining increasing attention for their inbuilt adjuvant capabilities. All these properties make nanoscale-type materials ideal reagents to be combined with structure-based, rational antigen and epitope design.

- **BOX 4: Epitope grafting approaches for multiple epitope presentation**

Most infections are characterized by diverse phases, for example acute and chronic phases, which require activation of both humoral and cell-mediated immune responses; thus, both B- and T-cell epitopes are vital components of protective vaccine formulations. Simultaneous, rather than consecutive presentation of epitopes is known to trigger stronger and more durable immune responses, therefore multiple epitopes should ideally be presented on the same polypeptide chain or via the same chemical scaffold. With regards to protein-based scaffolds, atomic level structure data and computational biology design methods can be used to remove an epitope from its host antigen and to transplant it onto a conformationally-apt region of an entirely different recipient protein

(Figure 5). This can endow the grafted epitope with new and improved immunological properties. For example, in its parental antigen, a neutralizing epitope may be masked by other less important, yet conformationally dominant epitopes; transplanting it to a new scaffold may improve its accessibility to the molecules governing the immune response. When the protein scaffold itself is an antigen, epitope grafting can be used to present multiple epitopes, *e.g*. T- and B-cell epitopes to trigger both the innate and humoral immune responses. In this case, attention must be made to ensure that the grafted epitope does not mask the host scaffold epitopes.

Figure Captions

Figure 1 – An overview of the different structure-based computational methods that are described in this review, starting from selection of conformational or linear epitopes originating, *e.g.*,from a viral capsid or from a protein antigen.

Figure 2 – A scheme of the different stages through which the Matrix of Low Coupling Energies method (MLCE) proceeds: from an antigen 3D structure and MD simulation, through the analysis of internal energetics and contacts, to prediction of uncoupled protein regions that can represent epitopes.

Figure 3 – A graphics representation of chemical methods that allow stabilizing an epitope conformation through cross-linking or 'stapling'.

Figure 4 – Broadly reactive antigens, able to trigger a neutralizing effect vs. multiple pathogenic strains, can be designed through the creation of chimeric proteins hosting mutations of variant-specific neutralizing epitopes in a single recombinant protein. The figure highlights the hot spots in the final product identified by Scarselli and coworkers in a broadly protective MenB antigen.

Figure 5 – Epitopes recognized within a given antigen can be transplanted onto a different antigen, to boost its immunogenic properties, provided that the conformational context in which the new epitope is to be grafted is compatible with its original structure.

Figure I, Box 3 – A schematic representation of the potential offered by decoration of nanoparticle or vesicles with peptide epitopes identified through the computational methods here described.

References

1 Thomas, S., *et al*. (2016) Future Challenges for Vaccinologists. *Methods Mol Biol* 1403, 41-55
2 Bloom, D.E., *et al*. (2017) Emerging infectious diseases: A proactive approach. *Proc Natl Acad Sci USA* 114, 4055-4059
3 Doran, T.M., *et al*. (2016) Chemical Tools To Monitor and Manipulate Adaptive Immune Responses. *J Am Chem Soc* 138, 6076-6094
4 Bethune, M.T. and Joglekar, A.V. (2017) Personalized T cell-mediated cancer immunotherapy: progress and challenges. *Curr Opin Biotechnol* 48, 142-152
5 Finn, O.J. (2008) Cancer Immunology. *N Engl J Med* 358, 2704-2715
6 Reynolds, P. and Marzi, A. (2017) Ebola and Marburg virus vaccines. *Virus Genes*, doi: 10.1007/s11262-11017-11455-x
7 McMichael, A.J. and Picker, L.J. (2017) Unusual antigen presentation offers new insight into HIV vaccine design. *Curr Opin Immunol* 46, 75-81
8 Skwarczynski, M. and Toth, I. (2016) Peptide-based synthetic vaccines. *Chem Sci* 7, 842-854
9 De Gregorio, E. and Rappuoli, R. (2014) From empiricism to rational design: a personal perspective of the evolution of vaccine development. *Nat. Rev. Immunol.* 14, 505-514
10 Rappuoli, R., *et al*. (2016) Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *The Journal of Experimental Medicine* 213, 469-481
11 Giuliani, M.M., *et al*. (2006) A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA* 103, 10834-10839
12 Altindis, E., *et al*. (2015) Protectome analysis: a new selective bioinformatics tool for bacterial vaccine candidate discovery. *Molecular & cellular proteomics: MCP* 14, 418-429
13 Malito, E., *et al*. (2015) Protein Crystallography in Vaccine Research and Development. *Int J Mol Sci* 16, 13106-13140
14 Liljeroos, L., *et al*. (2015) Structural and Computational Biology in the Design of Immunogenic Vaccine Antigens. *Journal of immunology research* 2015, 156241
15 Dormitzer, P.R., *et al*. (2012) Structural vaccinology starts to deliver. *Nature reviews. Microbiology* 10, 807-813
16 Scarselli, M., *et al*. (2011) Rational design of a meningococcal antigen inducing broad protective immunity. *Sci Transl Med* 3, 91ra62
17 Vita, R., *et al*. (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 43, D405-412
18 Toseland, C.P., *et al*. (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome research* 1, 4
19 Lefranc, M.P., *et al*. (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 37, D1006-1012
20 Ponomarenko, J., *et al*. (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *Bmc Bioinformatics* 9
21 Sun, J., *et al*. (2009) SEPPA: a computational server for spatial epitope prediction of protein antigens *Nucleic Acids Res* 37, W612-W616
22 Zhang, G.L., *et al*. (2005) MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res* 33, W172–W179
23 Zhang, L., *et al*. (2012) TEPITOPEpan: Extending TEPITOPE for Peptide Binding Prediction Covering over 700 HLA-DR Molecules. *PLoS ONE* 7, e30483.
24 Singh, H. and Raghava, G.P.S. (2001) ProPred: Prediction of HLA-DR binding

sites. . *Bioinformatics* 17, 1236-1237
25 Guan, J., *et al*. (2015) Structural comparison of four different antibodies interacting with human papillomavirus 16 and mechanisms of neutralization. *Virology* 483, 253–263
26 Ansari, H.R. and Raghava, G.P.S. (2010) Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome research* 6, https://doi.org/10.1186/1745-7580-1186-1186
27 Dalkas, G.A. and Rooman, M. (2017) SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. *BMC Bioinformatics* 18, 95
28 Peri, C., *et al*. (2015) Prediction of Antigenic B and T Cell Epitopes via Energy Decomposition Analysis: Description of the Web-Based Prediction Tool BEPPE. *Methods Mol Biol* 1348, 13-22
29 Scarabelli, G., *et al*. (2010) Predicting interaction sited from the energetics of isolated proteins: a new approach to epitope mapping. *Biophys*. *J*. 98, 1966-1975
30 Gourlay, L.J., *et al*. (2013) Exploiting the Burkholderia pseudomallei Acute Phase Antigen BPSL2765 for Structure-Based Epitope Discovery/Design in Structural Vaccinology. *Chemistry & Biology* 20, 1147-1156
31 Gourlay, L.J., *et al*. (2015) From crystal structure to in silico epitope discovery in the Burkholderia pseudomallei flagellar hook-associated protein FlgK. *Febs Journal* 282, 1319-1333
32 Peri, C., *et al*. (2016) Evolving serodiagnostics by rationally designed peptide arrays: the Burkholderia paradigm in Cystic Fibrosis. *Scientific Reports* 6, 24035
33 Soriani, M., *et al*. (2010) Exploiting antigenic diversity for vaccine design: the chlamydia ArtJ paradigm. *J*. *Biol*. *Chem*. 285, 30126-30138
34 Genoni, A., *et al*. (2012) Identification of Domains in Protein Structures from the Analysis of Intramolecular Interactions. *J*. *Phys*. *Chem*. *B*. 116, 3331-3343
35 Lassaux, P., *et al*. (2013) A structure-based strategy for epitope discovery in Burkholderia pseudomallei OppA antigen. *Structure* 21, 1-9
36 Fiorucci, S. and Zacharias, M. (2010) Prediction of protein-protein interaction sites using electrostatic desolvation profiles. *Biophys J* 98, 1921-1930
37 Lau, Y.H., *et al*. (2015) Peptide stapling techniques based on different macrocyclisation chemistries. *Chem Soc Rev* 44, 91-102
38 Walensky, L.D. and Bird, G.H. (2014) Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J Med Chem* 57, 6275–6288
39 Bird, G.H., *et al*. (2014) Stapled HIV-1 peptides recapitulate antigenic structures and engage broadly neutralizing antibodies. *Nat Struct Mol Biol* 21, 1058–1067
40 Gori, A., *et al*. (2016) Flexible vs Rigid Epitope Conformations for Diagnostic- and Vaccine-Oriented Applications: Novel Insights from the Burkholderia pseudomallei BPSL2765 Pa13 Epitope. *Acs Infectious Diseases* 2, 221-230
41 Buus, S., *et al*. (2012) High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Molecular & cellular proteomics: MCP* 11, 1790-1800
42 Carmona, S.J., *et al*. (2015) Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants. *Molecular & cellular proteomics: MCP* 14, 1871-1884
43 Lu, Y., *et al*. (2015) Chimeric peptide constructs comprising linear B-cell epitopes: application to the serodiagnosis of infectious diseases. *Sci Rep* 5, 13364
44 Gearhart, T.L., *et al*. (2016) Selection of a potential diagnostic biomarker for HIV

infection from a random library of non-biological synthetic peptoid oligomers. *Journal of immunological methods* 435, 85-89
45 Bovier, P.A. (2008) Epaxal®: a virosomal vaccine to prevent hepatitis A infection. *Expert Review of Vaccines* 7, 1141–1150
46 Nakanishi, T., *et al*. (1999) Positively charged liposome functions as an efficient immunoadjuvant in inducing cell-mediated immune response to soluble proteins. *J Control Release* 61, 233-240
47 Henriksen-Lacey, M., *et al*. (2011) Liposomal vaccine delivery systems. *Expert Opinion on Drug Delivery* 8, 505–519
48 Kitamura, H., *et al*. (1997) CD1d-restricted and TCR-mediated activation of Vα14 NKT cells by glycosylceramides. *Science* 278, 1626-1629
49 Schwendener, R.A. (2014) Liposomes as vaccine delivery systems: a review of the recent advances. *Therapeutic Advances in Vaccines* 2, 159–182
50 Bachmann, M.F. and Jennings, G.T. (2010) Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol* 10, 787–796
51 Harding, C.V., *et al*. (1991) Liposome-encapsulated antigens are processed in lysosomes, recycled, and presented to T cells. *Cell* 64, 393–401
52 Irvine, D.J., Hanson, M. C., Rakhra, K., & Tokatlian, T. (2015). Synthetic Nanoparticles for Vaccines and Immunotherapy. Chemical Reviews, 115(19), 11109–11146. (2015) Synthetic Nanoparticles for Vaccines and Immunotherapy. *Chemical Reviews* 115, 11109–11146
53 Hanson, M.C., *et al*. (2015) Liposomal vaccines incorporating molecular adjuvants and intrastructural T-cell help promote the immunogenicity of HIV membrane-proximal external region peptides. *Vaccine* 33, 861-868
54 Torres, A.G., *et al*. (2015) Protection of non-human primates against glanders with a gold nanoparticle glycoconjugate vaccine. *Vaccine* 33, 686-692
55 Serruto, D., *et al*. (2012) The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. *Vaccine* 30 Suppl 2, B87-97
56 Aagaard, C., *et al*. (2011) A multistage tuberculosis vaccine that confers efficient protection before and after exposure. *Nature medicine* 17, 189-194
57 McLellan, J.S., *et al*. (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* 480, 336-U386
58 Capelli, R., *et al*. (2017) SAGE: A Fast Computational Tool for Linear Epitope Grafting onto a Foreign Protein Scaffold. *J. Chem. Inf. Model*. http://dx.doi.org/10.1021/acsinfecdis.5b00118
59 Pizza, M., *et al*. (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287, 1816-1820
60 Borrebaek, C.A.K. and Wingren, C. (2009) Design of high-density antibody microarrays for disease proteomics: Key technological issues. *J. of Proteomics* 2009, 928-935
61 Cretich, M., *et al*. (2014) Protein microarrays technology: how far off is routine diagnostics? . *Analyst* 139 528-542
62 Vetro, M., *et al*. (2017) Preparation and immunogenicity of gold glyco-nanoparticles as antipneumococcal vaccine model. *Nanomedicine* 12, 13-23
63 Ferrari, G., *et al*. (2006) Outer membrane vesicles from group B Neisseria meningitidis Δgna33 mutant: Proteomic and immunological comparison with detergent-derived outer membrane vesicles. *Proteomics* 6, 1856-1866
64 Bartolini, E., *et al*. (2013) Recombinant outer membrane vesicles carrying Chlamydia muridarum HtrA induce antibodies that neutralize chlamydial infection in

vitro. *Journal of Extracellular Vesicles* 2, 10.3402/jev.v3402i3400.20181
65 Fantappiè, L., *et al*. (2014) Antibody-mediated immunity induced by engineered Escherichia coli OMVs carrying heterologous antigens in their lumen. *Journal of Extracellular Vesicles* 3, 10.3402/jev.v3403.24015

Outstanding Questions Box

- Technological advances in computing are steadily growing, while hardware is getting cheaper and cheaper: can this turn into *in silico* analyses that help predict antigens and epitopes more directly/accurately?
- Given the increasing number of 3D structures provided by X-ray crystallography, there will be an increasing possibility of generating homology models for newly discovered antigens. Will these modeled structures allow accurate epitope prediction?
- Will computational methods be able carry out large-scale predictions on all available antigen structures, and to generate peptide libraries for immediate screening for diagnostic/therapeutic potential?
- Will we be able to formulate basic sequence/structural principles and integrate them into existing *in silico* epitope prediction methods to take predictions one step further and discriminate protective from non-protective epitopes, ultimately predicting vaccinological protection in humans?
- Will the above mentioned approaches be combined in methods that integrate big data from disparate sources (structural, clinical, epidemiological) as the basis for machine learning, artificial intelligence approaches aimed at predicting immunologic potential from sequence/structure information?
- Will epitope-grafted antigens presenting multiple T-/B-cell epitopes induce better protection with regards to their single counterparts?
- Can optimal carriers for epitope presentation be designed rationally and immediately applied to disparate needs?
- Will it be possible to overcome current limitations in diagnostic accuracy using multiple highly-selective epitopes as probes?
- Will epitope-based diagnostic solutions be able to use a single drop of blood or saliva?

Figure 1

Click here to download Figure Figure_1-Figure_300dpi_5inch_version3.tif

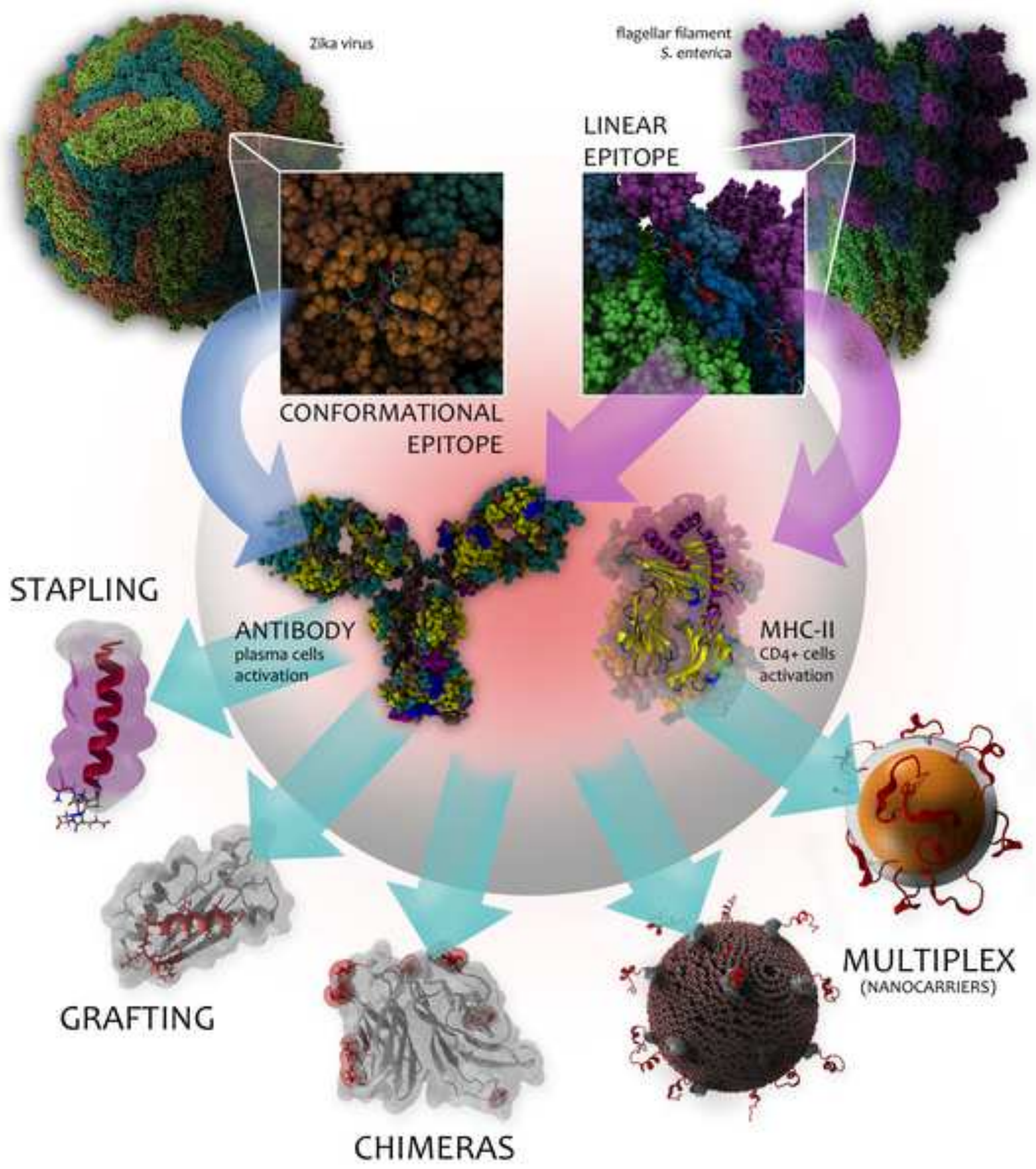

Figure 2

Click here to download Figure Figure_2-MLCE.tif

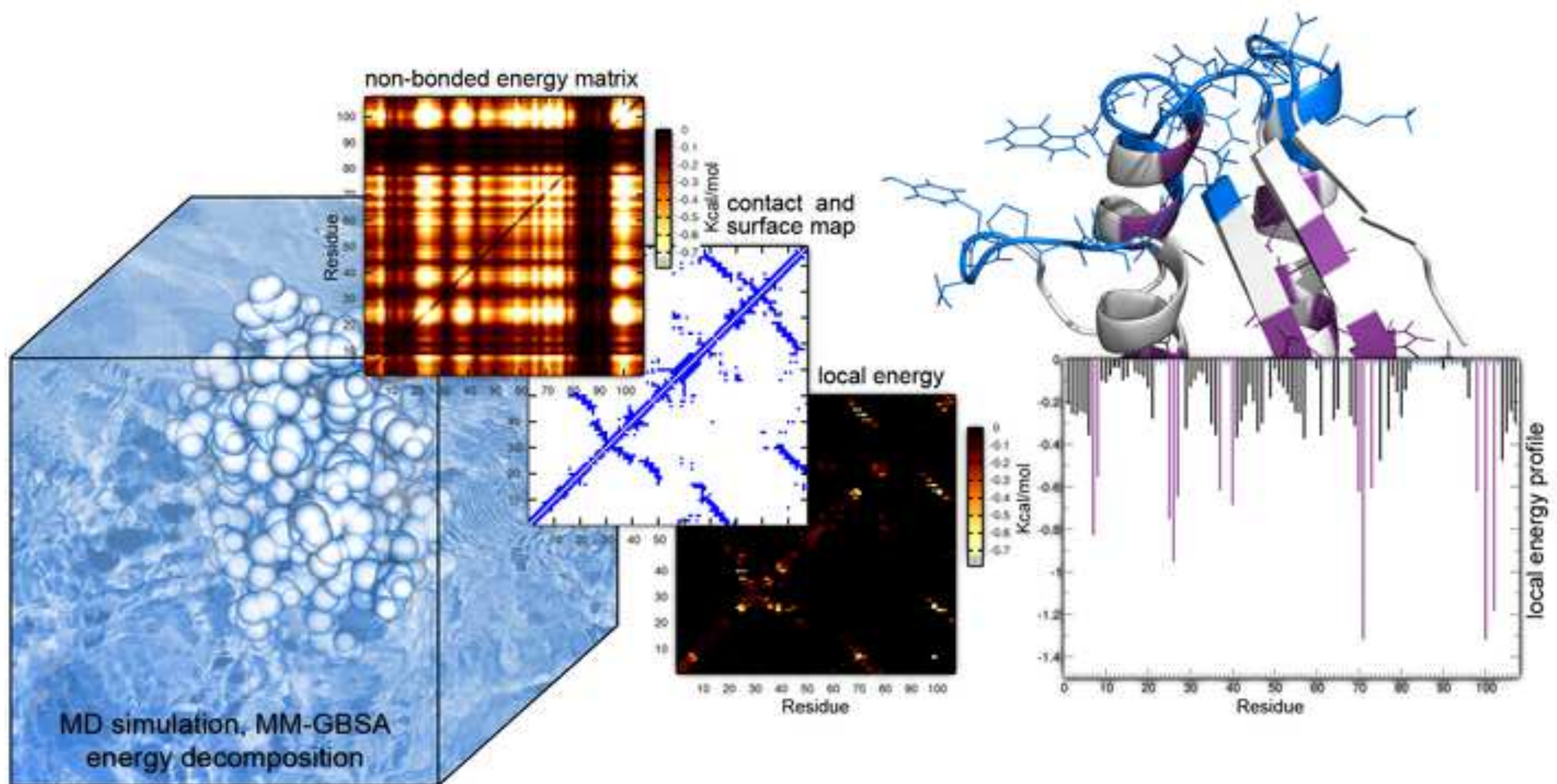

Figure 3

Click here to download Figure Figure_3-stapling3_300dpi.tif

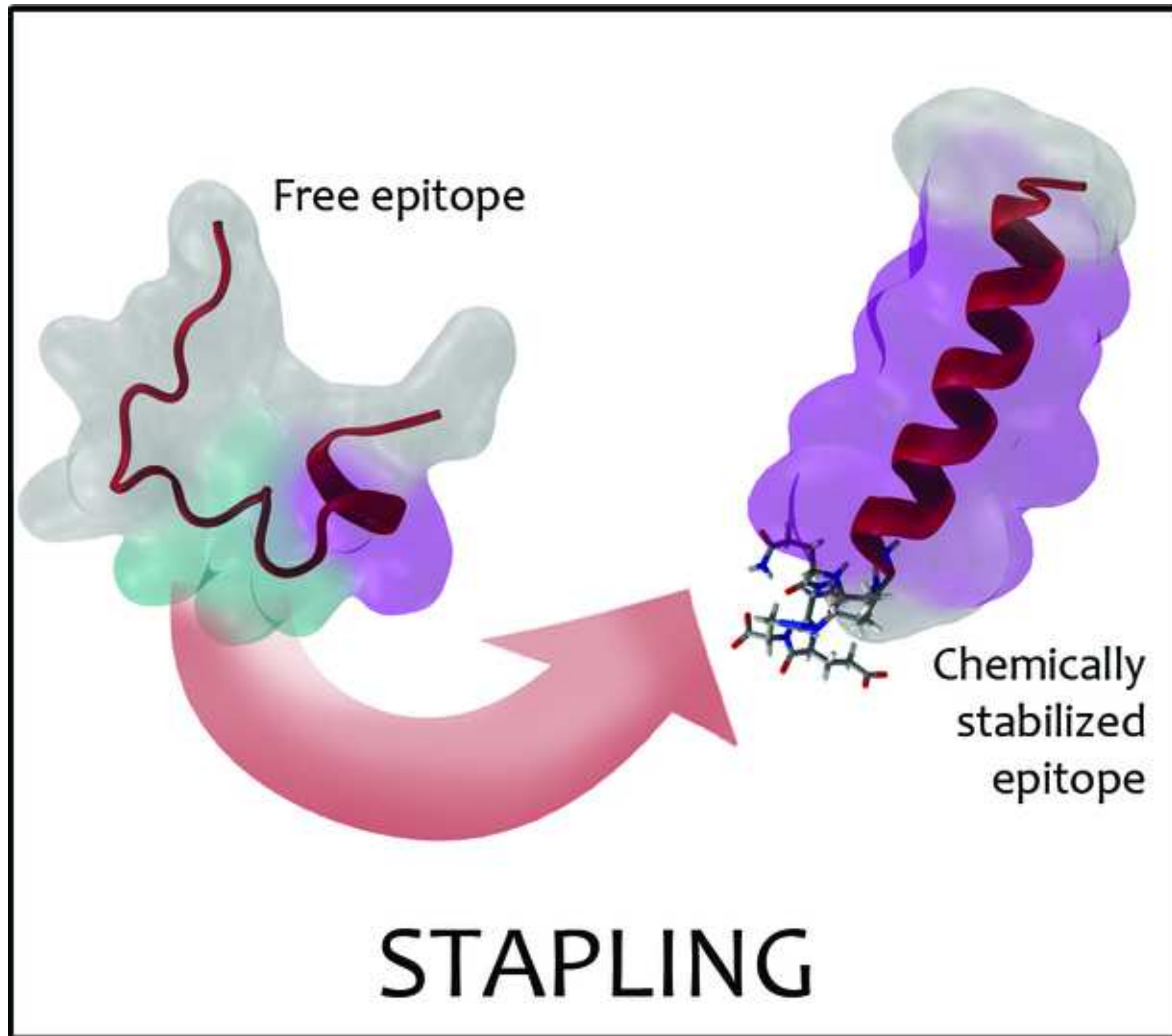

Figure 4

Click here to download Figure Figure_4-chimeras2_300dpi.tif

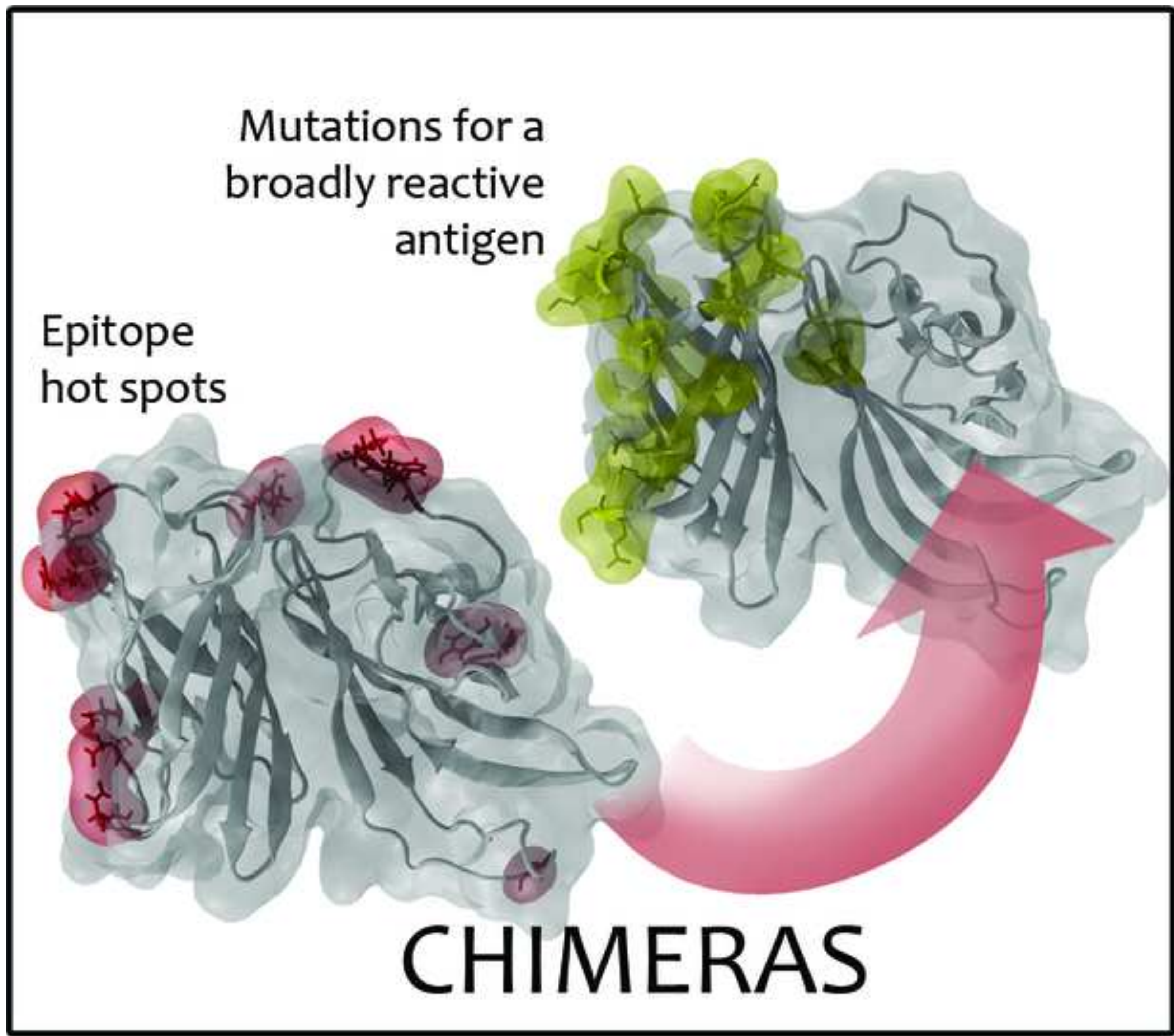

Figure 5

Click here to download Figure Figure_5-grafting3_300dpi.tif

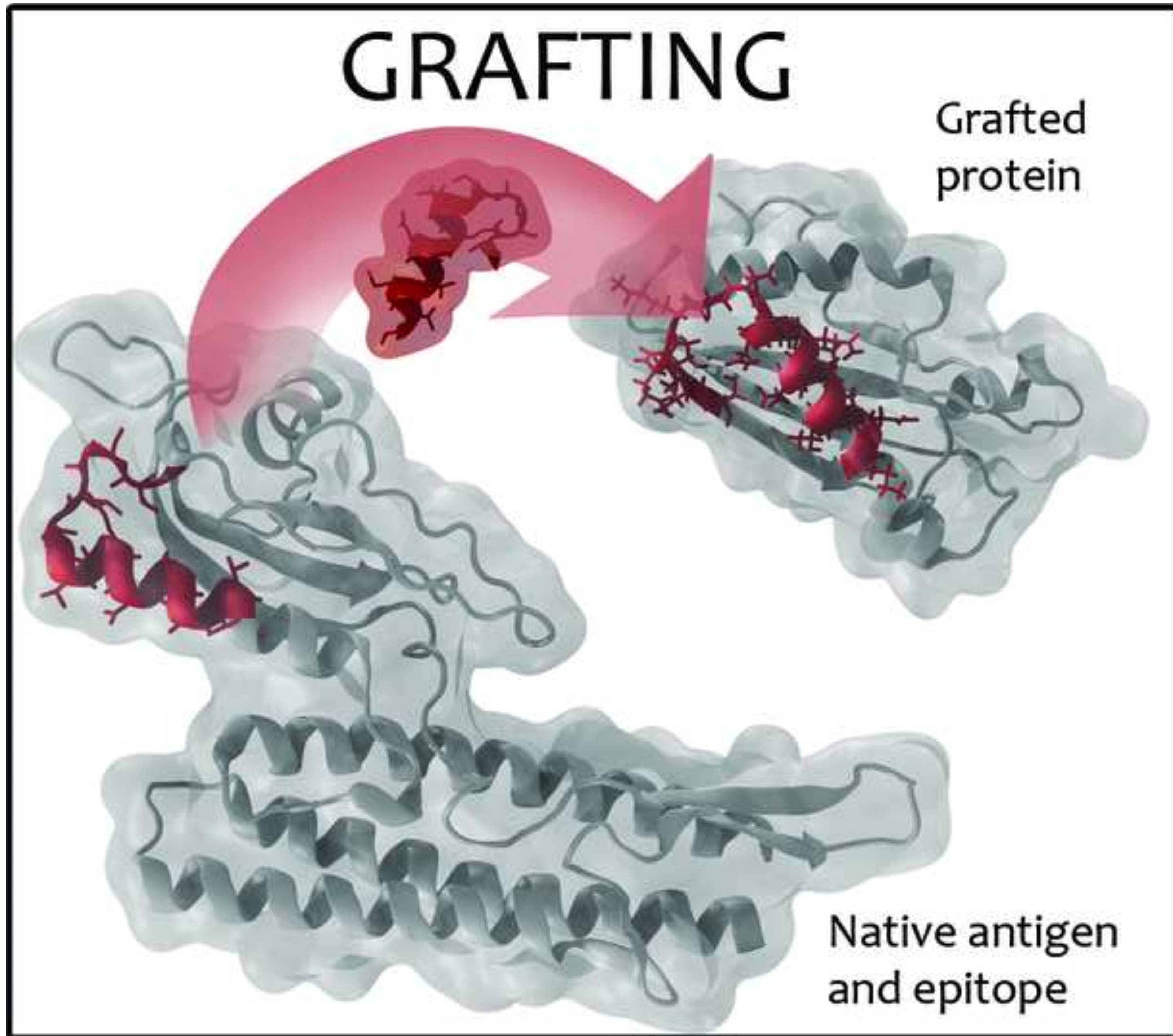

Figure I, BOX 3

Click here to download Figure Figure_I-BOX3-nanocarr-multiplex_300dpi.tif