15 Predicting and Dissecting High-order Molecular Complexity by Information-driven Biomolecular Docking

Panagiotis L. Kastritis and Alexandre M.J.J. Bonvin
Bijvoet Center for Biomolecular Research, Science Faculty, Utrecht University, Utrecht, The Netherlands

15.1 Increased Molecular Flexibility and Complexity in Antimicrobial Drug Design

Drug discovery today is moving towards the development of more complicated agents, especially in the field of antimicrobial drug design. In order to enhance lead compounds potency and optimize the design of more successful lead molecules, the incorporation of structural knowledge is deemed necessary (Hajduk and Greer, 2007). This process is called structure-based drug design (SBDD), the process of finding new medications based on the knowledge of the structure and function of the biological target of interest, generally by using computer modelling (docking).

Molecular docking is a computational method used to predict the preferred binding mode of one molecule to another, starting from their unbound conformations. In protein–ligand docking, the candidate target is usually a protein with medical relevance, and the lead compound is (usually) an organic compound that can inhibit the activity of the protein (see Chapters 3–5, this volume). SBDD has already demonstrated its power through the discovery of novel therapeutics over the years (Simmons *et al.*, 2010). For example, knowledge of the human immunodeficiency virus (HIV) protease three-dimensional structure enabled the

design and optimization of five inhibitors that are now commercially available antiretroviral drugs (Erickson *et al.*, 1990; Roberts *et al.*, 1990; Dorsey *et al.*, 1994). SBDD has been successfully applied to the development of other drugs, including zanamivir for influenza (GlaxoSmithKline; McCauley, 1999) and several non-steroidal anti-inflammatory agents targeting cyclooxygenase 2 (e.g. celecoxib, Pfizer; Stratton and Alberts, 2002). New approaches in SBDD, such as fragment-based drug design (FBDD) continue to flourish, opening the route to the *ab initio* design of agents with increased ligand potency (Murray and Blundell, 2010). FBDD is based on the idea that the use of small inhibitors as building blocks can lead to the development of larger compounds with higher affinity. FBDD is moving towards the design of more complex and affine biomolecules. Analysis and understanding of the increased complexity and flexibility of such ligands is deemed necessary for success in drug design. Current developments in antimicrobial agents are not limited to SBDD/FBDD methods; there has also been special interest in the discovery, development and application of (natural) complex organic compounds (Cowan, 1999; Hann *et al.*, 2001), hybrid molecules derived from click chemistry (Kolb and Sharpless, 2003) and peptide antimicrobial agents (Jenssen *et al.*, 2006)

(see Fig. 15.1). Click chemistry encompasses all chemical procedures tailored to generate molecules quickly and reliably by joining small units together. The latter is inspired by the fact that nature also generates substances by joining small modular units, such as proteins, sugars, nucleic acids and lipids.

A novel approach in modern drug discovery, which challenges the traditional concept of inhibiting enzymes, is to design inhibitors that specifically target protein–protein complexes, blocking or modulating their underlying interactions (Arkin and Wells, 2004). Indeed, protein–protein interactions should also form an important class of therapeutic targets (Patel and Player, 2008) as they are involved in nearly all normal and pathological pathways, including signal transduction, cell adhesion, proliferation, growth, differentiation, structure of cellular cytoskeleton, programmed cell death and virus self-assembly (Toogood, 2002). Inhibitors designed for this category of targets are usually larger and more complex compared with traditional ligands that target enzymes. The interface that these ligands have to inhibit is also larger compared with an enzymatic active site. A collection of crystallographically determined protein–protein complexes with known inhibitors has been compiled in the 2P2I protein–protein inhibition database (http://2p2idb.cnrs-mrs.fr/; Bourgeas *et al.*, 2010). As of February 2012, we have calculated that the mean molecular weight of all 45 ligands present in the database, targeting two classes of protein–protein interactions and 10 different protein–protein complexes, is 535±156 Da.
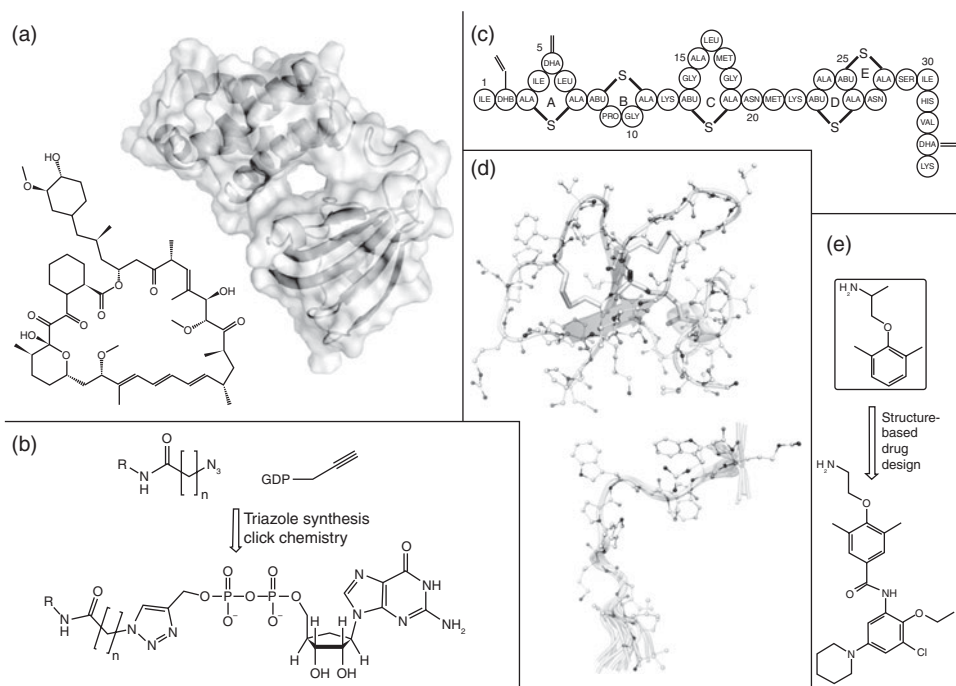
Fig. 15.1. Complexity of molecules used to inhibit protein molecules. (a) The protein–protein complex of immunophilin–immunosuppressant FKBP12 is inhibited by rapamycin (PDB ID: 1FAP). (b) Triazole synthesis used in click chemistry in order to derive molecules of higher-order molecular complexity. Azide fragments are coupled with an alkyne GDP core. (c) The nisin A molecule composed of natural and unnatural amino acids, such as didehydroalanine (DHA), didehydroaminobutyric acid (DHB) and α-aminobutyric acid (ABU). (d) Natural peptide inhibitors, human β-defensin-2 (PDB ID: 1FQQ) and indolicidin, an unstructured peptide (PDB ID: 1G89). (e) Process of structure-based drug design, where a new, larger urokinase inhibitor with higher potency is synthesized (50% inhibitory concentration (IC_{50}) = 0.003 μM) based on the initial inhibitor (IC_{50} = 0.91 μM).

All newly emerging methods for designing potential antimicrobial agents (Fig. 15.1) have to deal with a common challenge: the increased complexity of the complexes that involve larger and rather complicated molecules. These are typically rather flexible systems, behaving more like fluids than rigid bodies, in contrast to what is usually assumed in SBDD, where proteins are treated as rigid entities (mostly due to the computational cost, as very large libraries of compounds have to be screened against a target receptor molecule). For a recent review on protein flexibility and its role in drug design, see Fuentes *et al*. (2011).

Since the discovery of the importance of flexibility in biomolecular association, for example with the flexible protein recognition model of Grunberg *et al*. (2004) where recognition is proposed to occur in sequential steps involving: (i) diffusion; (ii) conformer selection from a pool of conformers; and (iii) induced fit, some progress has been achieved in the inclusion of flexibility in SBDD (Meagher and Carlson, 2004). Nevertheless, these are rather limited, as tackling both receptor and ligand flexibility simultaneously with conventional methods remains very challenging due to the explosion in the number of degrees of freedom of the system under study, which can translate into exorbitant computational costs.

A multipurpose docking program that can be used to address this type of challenge is HADDOCK (high-ambiguity-driven biomolecular docking) (Dominguez *et al*., 2003; de Vries *et al*., 2007). In this chapter, we will describe this program and its successful applications to protein–small ligand docking, with specific emphasis on antimicrobial agents, and will consider this with regard to the treatment of flexibility and its associated challenges. In Section 15.2, the key idea fundamental to data-driven biomolecular docking is introduced: how can a wide variety of experimental and/or predicted information be used to drive the modelling process, thereby restricting the interaction space to be searched? Relevant technical and theoretical aspects are discussed in this section, where treatment of molecular complexity and flexibility is introduced. In this way, the reader should be able to understand better the subsequent sections of this chapter, where applications of data-driven docking in small-molecule design are described. Specifically, in Section 15.3, different examples from recent literature are portrayed in which data-driven docking has been successful in providing structural insights for the design of small-molecule inhibitors that can act as antimicrobial agents.

15.2   Data-driven Docking

X-ray crystallography and nuclear magnetic resonance (NMR) are widely used experimental techniques to obtain atomic resolution structures of protein–ligand complexes and unravel the structural details of the recognition process. However, their traditional use often implies high costs in terms of time, resources and maintenance, whereas their applicability to protein–ligand complexes is strongly case dependent (Jahnke, 2007). On the other hand, advances in X-ray crystallography and NMR have expanded the range of tractable targets along with improving the overall throughput (Blundell *et al*., 2002; Betz *et al*., 2006). For example, if the X-ray structure of the target is known, a simple crystallographic screening may be used. Ideally, the active site of the target macromolecule should be open to solvent channels in the crystal to allow complex formation by the ligand-soaking method: the crystal is soaked in a solution containing a mixture of compounds; from this, the most potent ligand will bind in the active site of the crystalline macromolecule by diffusion into the crystal. X-ray crystallography might, however, fail to yield a protein–ligand structure for various reasons, often because some proteins simply do not crystallize. Other possible reasons for failure include the fact that the ligand molecule might occupy the active site of the enzyme within the crystal insufficiently or in a disordered manner that makes the electron density much less defined, or because crystal packing might prevent binding by the fragments. A case representing the abovementioned crystallographic artefact can be found in the structure of the homodimeric enzyme malate dehydrogenase, which was

crystallized as a tetramer (Protein Data Bank (PDB) ID: 4MDH; http://www.rcsb.org/pdb/). In 4MDH, the chains building the homodimer are involved in different crystal contacts: Chain A has crystal contacts near the catalytic site, whereas chain B does not show such contacts. As a consequence, the conformation of loop 89–104 close to the active site is extensively affected. In one of the monomers, this loop is surrounded by solvent while in the other it contacts another chain in the crystal. This results in a different conformation of the loop, which in turn affects the conformation of the active site. Another possible problem with crystallographic studies is that ligand binding is often achieved by soaking crystals of the protein into a solution containing the ligand. For large and flexible ligands, this soaking procedure might disrupt the crystal.

NMR can provide a powerful alternative to X-ray crystallography as it allows studying of the binding and behaviour of molecules in solution and can provide three-dimensional structural information about the complex. Developments in the field of NMR are therefore of particular interest, as a vast amount of experimental data relevant to the protein–ligand system can be extracted, provided that the structure of the target is known. For example, depending on the size of the target, the labelling schemes and the binding regime, a variety of NMR data can be obtained relevant to the ligand and/or the protein sides, such as chemical shift perturbation (CSP) data, nuclear Overhauser effects (NOEs), residual dipolar couplings and cross-correlation rates. Data sources that can provide information about a complex are, however, not restricted to NMR. For example, even simple mutagenesis experiments can provide valuable information that can assist a modelling procedure. All these data can be used in sophisticated algorithms to model protein–ligand complexes *in silico*. Such algorithms are referred to as information-driven (data-driven) and rely on such data to derive three-dimensional models of the systems in atomistic detail. A unique computational method falling under this category is HADDOCK (Dominguez *et al.*, 2003).

HADDOCK is an information-driven flexible docking approach for the modelling of biomolecular complexes (Dominguez *et al.*, 2003). Compared with other docking methods, HADDOCK is unique in the sense that it can handle a wide variety of experimental and/or bioinformatics data to drive the modelling process (Melquiond and Bonvin, 2010). The method allows for a rather sophisticated treatment of flexibility by limiting the search to the relevant interaction space of the biomolecules that are being docked. The program incorporates information about the interface regions of the binding molecules (a binding pocket/active site of an enzyme is also considered an interface). The latter can be identified by several experimental methods/techniques, including mutagenesis in combination with a binding assay, chemical modifications (e.g. by cross-linkers or oxidative agents) detected by mass spectrometry (MS), hydrogen/deuterium exchange detected by either MS or NMR, and a variety of valuable NMR data such as CSP, cross-saturation transfer, INPHARMA (protein-mediated interligand NOEs for pharmacophore mapping; Sanchez-Pedregal *et al.*, 2005) and structure–activity relationships by interligand NOEs (Becattini and Pellecchia, 2006). Bioinformatics predictions, for example based on evolutionary information, can also be used when experimental data are scarce or unavailable (de Vries and Bonvin, 2008). As well as being able to deal with such a large variety of experimental and/or predicted information, HADDOCK also supports classical NMR restraints such as distances from NOEs and paramagnetic relaxation enhancement measurements, dihedral angles, residual dipolar couplings, diffusion anisotropy restraints and pseudo-contact shifts, the latter three providing valuable information about the relative orientation of the components in a complex. For more information about useful sources for restraining the docking, see Melquiond and Bonvin (2010), Schmitz *et al.* (2012) and van Dijk *et al.* (2005).

Most of the information sources described in the previous paragraph typically only identify or predict interfacial regions, and do not define the contacts across an interface. In HADDOCK, these are implemented as ambiguous interaction restraints (AIRs)

that will force the interfaces to come together without imposing a particular orientation. AIRs are entered as a list of active and passive residues. The active residues correspond to the (experimentally) identified interface residues, whereas passive residues correspond to their solvent-accessible neighbouring residues. The latter ensure that residues located in the interface but not detected can satisfy the AIRs. Note that this terminology is not restricted to amino acid residues of proteins, despite the fact that the algorithm was originally developed for protein–protein docking; for example, a residue can also be a non-standard amino acid, a nucleotide base, a sugar or any organic compound. An AIR corresponds to an ambiguous intermolecular distance (d_{iAB}^{eff}) with a maximum value of typically 2 Å between any atom m of an active residue i of protein A (m_{iA}) and any atom n of both active and passive residues k (N_{resB} in total) of protein B (n_{kB}) (and inversely for protein A). The effective distance, corresponding to each restraint is calculated using the following equation:

$$d_{iAB}^{eff} = \left(\sum_{m_{iA}=1}^{N_{Aatoms}} \sum_{k=1}^{N_{resB}} \sum_{n_{kB}=1}^{N_{Batoms}} \frac{1}{d_{m_{iA}n_{kB}}^{6}} \right)^{-\frac{1}{6}} \qquad (15.1)$$

where $\frac{1}{d_{m_{iA}n_{kB}}^{6}}$ denotes a potential that resembles the Lennard–Jones attractive term.

The function has the property that all d_{iAB}^{eff} will always be smaller than the shorter distance $d_{m_{iA}n_{kB}}$ entering the sum. The AIRs effectively enforce the defined interfaces to come together without imposing any restraint on their relative orientation. The AIRs can be further fine-tuned manually to restrict them to specific atoms or groups of atoms (for example, pharmacophore groups).

15.2.1 Dealing with molecular flexibility

HADDOCK deals with molecular flexibility at different levels, both implicitly and explicitly, starting from the initial coordinate files of the biomolecules until the final flexible refinement in explicit water with subsequent energy minimization.

First level – implicit (as starting point for the docking)

ENSEMBLE DOCKING. The HADDOCK program supports ensemble docking, meaning that it can handle as input more than one configuration of any of the partners, such as, for example, an ensemble of NMR structures or different crystal structures of the same enzyme. Other methods can also be used to produce ensemble of structures for docking, such as molecular dynamics simulation, normal modes analysis and principal components analysis; however, these will not be covered in this chapter.

In such cases, HADDOCK performs a cross-docking of all possible combinations of starting structures. Ideally, the number of initial rigid-body docking poses generated should be a multiple of the number of all combinations, with each combination sampled multiple times (e.g. at least 100). Ensembles that are too large might result in a dilution effect, meaning by that, if only a few structures have a proper conformation for binding, only a small fraction of all sampled combinations might lead to a successful docking.

FRAGMENT-BASED DOCKING. If the target protein molecule undergoes significant conformational changes, a multibody docking protocol can be used (Karaca et al., 2011). The protein target is then treated as a collection of separate domains (Karaca and Bonvin, 2011).

TREATMENT OF ENZYME FLEXIBILITY IN HADDOCK. A good example of flexibility treatment is the case of the enzyme dihydrofolate reductase (DHFR). The enzyme catalyses the reduction of 5,6-dihydrofolate to 5,6,7,8-tetrahydrofolate, utilizing NADPH as a cofactor (Sawaya and Kraut, 1997). DHFR is important in drug discovery, as blockade of its enzymatic activity leads to irreversible cell death. A large number of crystal structures of the enzyme complexed with different compounds and substrates have been deposited in the PDB. For example, if docking of a specific compound and Escherichia coli DHFR is performed, one can use all 55 experimental structures of the E. coli enzyme in the PDB (as of February 2012) (a requisite is, however,

that they are all consistent with each other and contain the same atoms – missing fragments/side chains should thus be added prior to docking). The ensemble docking method allows simultaneous docking of the compound to all chains, implicitly treating the receptor flexibility in the initial rigid-body stage of the docking by the collection of different conformers. Similarly, different conformations of the second molecule can be introduced (for example, one more), leading to 2 × 55 different combinations of starting structures for the docking.

The catalytic mechanism of the *E. coli* DHFR enzyme was deciphered by Sawaya and Kraut (1997), who concluded that the M20 loop adopts different configurations when the substrate and the cofactor of the enzyme are bound. Therefore, DHFR can be treated as a collection of two domains, one corresponding to the enzyme without the loop and a second corresponding to the M20 loop. During a docking run with a lead compound, a simultaneous three-body docking protocol can be applied, connecting the M20 loop at the hinge regions with the rest of the molecule by defining additional distance restraints. Such treatment of the system might allow discovery of a possible low-energy orientation of both the M20 loop and the lead compound in the active site of the enzyme simultaneously.

Second level – explicit (during docking)

The docking protocol in HADDOCK, which makes use of the crystallography and NMR system package (Brunger *et al.*, 1998) as computational engine, consists of three successive steps:

1. Rigid-body energy minimization. At this step, called *it0*, the molecules are brought together as rigid units by energy minimization, using the effective distance criterion (see above) and non-bonded energies (electrostatic and van der Waals energies) that become effective once the molecules are within the non-bonded cut-off (typically 8.5 Å).

2. Semi-flexible refinement in torsion angle space. Typically, the top 10–20% of the models in *it0* are subsequently subjected to flexible refinement in torsion angle space (the *it1* step). Selection of the structures

to be subjected to this refinement stage is based on the HADDOCK score, which is a weighted sum of various terms (buried surface area, empirical desolvation term, electrostatic, van der Waals and restraint violation energies) (see de Vries *et al.*, 2007 for details). The flexible regions are by default automatically defined as all residues within 5 Å of the partner molecule plus their preceding sequential neighbour. A manual definition of flexible segments is also possible. The flexible refinement stage consists of three simulated annealing refinements and a final steepest-descent energy minimization:

- In the first simulated annealing stage, the molecules are treated as rigid entities and their relative orientation is optimized.
- In the second stage, side chains at the interface are allowed to move to optimize the packing. In the case of protein–ligand docking, the ligand molecule can be treated as fully flexible (see below).
- In the third stage, both side chains and backbone at the interface are allowed to move to allow for small conformational rearrangements.
- A final energy minimization is applied in order to optimize the derived complexes and calculate the underlying energetics for each docked solution that derived from *it1*.

3. Final refinement in explicit solvent. The final stage consists of a gentle refinement in explicit water, or in DMSO for hydrophobic molecules (e.g. transmembrane proteins). The system is first heated to 300 K with position restraints on all atoms except for the flexible side chains at the interface. Short molecular dynamics simulation steps are then performed at 300 K with position restraints only on non-interface backbone-heavy atoms. During the final cooling stage (reaching finally 100 K), the position restraints are limited to backbone atoms outside the interface. A final energy minimization step is again applied and the final energy of each derived complex is calculated.

Only by an explicit treatment of flexibility during the refinement stages of the docking can well-packed biomolecular models

be obtained. These typically do not contain intermolecular clashes, resembling in this aspect structures deposited in the PDB.

Note that one can define *fully flexible segments* that are treated as fully flexible throughout the entire docking run (except the initial rigid-body minimization). This is, for example, one way of increasing the sampling of conformations in case of very flexible ligands or unstructured peptides.

15.2.2   Ways of addressing molecular complexity within data-driven docking

HADDOCK was developed initially for the prediction of protein–protein complexes (Dominguez *et al.*, 2003). Since its initial implementation, HADDOCK has extended its functionalities to account for a variety of molecules, including nucleic acids, sugars and small molecules. The program also supports a number of modified amino acid residues and inorganic elements. With its solvated docking implementation (Kastritis *et al.*, 2012; van Dijk and Bonvin, 2006), docking of fully solvated molecules can be performed. The HADDOCK web server (http://haddock.chem.uu.nl/services/HADDOCK/; further information on manuals and tutorials is given at the end of the chapter; de Vries *et al.*, 2010b) offers a user-friendly interface. When a PDB file of a small molecule is uploaded, topology and parameter files of the ligand are automatically retrieved from the PRODRG server (http://davapc1.bioch.dundee.ac.uk/prodrg/; Schuttelkopf and van Aalten, 2004). Additionally, protonation states of histidine residues are also considered by the program, and when not supplied by the user, these are assigned automatically using the WHAT IF web server (http://swift.cmbi.ru.nl/servers/html/; Vriend, 1990).

Another important feature is that the code is not restricted to bivalent molecular docking. The program supports up to six-body docking, meaning that up to six (bio)molecules can be docked simultaneously, independent of their molecular type. HADDOCK is therefore suitable for drug design targeting biomolecular complexes, as these systems might be composed of more than two molecules or domains.

15.3   Applications of Data-driven Docking in Antimicrobial Drug Discovery and Beyond: When Flexibility Matters

Data-driven docking has been applied extensively to a large variety of systems and has shown a very strong performance in the blind critical assessment of the prediction of interactions (see CAPRI (critical assessment of predicted interactions), a community-wide experiment on the comparative evaluation of protein–protein docking for structure prediction: http://www.ebi.ac.uk/msd-srv/capri/; de Vries *et al.*, 2010a; Lensink and Wodak, 2010). A considerable number of experimental structures of complexes calculated using HADDOCK have been deposited into the PDB, including structures of proteins with antibacterial activity, such as hydramycin-1 (Jung *et al.*, 2009).

Data-driven docking can also be used to model protein–ligand interactions and give structural insights and details of enzymatic or inhibitory mechanisms. For this, however, a modified version of the default HADDOCK protocol should preferably be used: during all docking steps, the ligand must be kept fully flexible while the residues within the protein-binding site are only defined as active for the rigid-body docking stage (*it0*) and considered as passive for the subsequent semi-flexible refinement stage (*it1*). This strategy effectively pulls the ligand within the binding site during rigid-body docking while allowing a more thorough exploration of the binding pocket during the refinement stage.

At the end of the protocol, clustering based on pairwise root mean square deviation criteria is performed and the lowest energy structure of the lowest energy cluster is usually taken as the best solution, depending on the quality of the experimental data available. Clustering is automatically performed by HADDOCK and is an integral part of the docking and scoring procedure. It is, however, recommended to check a number

of (top) clusters and their statistics and also inspect their three-dimensional structures to comprehend the docking predictions. If possible, additional information might be used, if available, to guide the selection. For clustering of protein–ligand solutions, it is recommended to lower the clustering cut-off from the default value of 7.5Å to a small value (e.g. 2.0Å). Protein–ligand docking options are easily accessible via the web server version of HADDOCK (de Vries *et al*., 2010b), under the guru interface. Note that this interface is meant for experienced users only (for details on the HADDOCK protocol and its web server, see de Vries *et al*., 2010b; Melquiond and Bonvin, 2010; Schmitz *et al*., 2012). In the future, a specific protein–ligand interface will be made available, but at this time a ligand should be entered as 'protein' in the web interface using the HETATM fields in the PDB file.

Applications of data-driven docking on protein–ligand systems are extensive (Rutten *et al*., 2006; Song *et al*., 2007; Tomaselli *et al*., 2007; Wu *et al*., 2007; Arnusch *et al*., 2008; Rutten *et al*., 2009; Krzeminski *et al*., 2010; Schneider *et al*., 2010; Fiamegos *et al*., 2011), including the dissection of catalytic mechanisms concerning several membrane and soluble enzymes, as well as the characterization of inhibitors/antimicrobial agents for targets of medical relevance. Rather flexible ligands have also been studied, skipping the initial rigid-body search step and performing the docking in a fully flexible manner during the refinement stage. This has been applied successfully to decipher structure–activity relationships between lectins and oligosaccharides (Wu *et al*., 2007). Animation movies corresponding to applications of the two different protein–ligand docking protocols, illustrating the docking procedure and the interaction between lectins and oligosaccharides can be found in the following links:

- standard protein–ligand protocol (Krzeminski *et al*., 2010): http://www.nmr.chem.uu.nl/haddock/movies/hgal3.html.
- fully flexible docking, skipping the initial rigid-body docking (Wu *et al*., 2007): http://www.nmr.chem.uu.nl/haddock/movies/cg1.html.

15.3.1 Unveiling substrate specificity and catalytic mechanisms of outer-membrane proteins from pathogenic Gram-negative bacteria using HADDOCK

The lipid A portion of lipopolysaccharide is the major component of the outer leaflet of the outer membrane of Gram-negative bacteria, which is toxic to humans. The toxicity of lipid A can be reduced by modifications, often accomplished by specific enzymes located in the same cellular compartment. For these enzymes, data-driven docking has provided useful insights into their catalytic mechanisms (Rutten *et al*., 2006, 2009), thereby opening the route for the SBDD of novel antimicrobial agents for these targets.

PagL from *Pseudomonas aeruginosa* is an outer-membrane lipid A deacylase located in the bacterial cell wall. PagL hydrolyses the ester bond at the 3-position of lipid A, thereby releasing the primary 3-OH C14 moiety. Due to the protein's localization and because lipid A is modified by PagL, this protein can help bacteria to finally evade the host immune system. Therefore, PagL is a suitable target for antimicrobial agents; however, its structure and catalytic mechanisms, which were still unknown at the time, were unveiled in this study (Rutten *et al*., 2006). Interaction restraints between lipid A and PagL were derived based on the following information: (i) the crystallized protein is active; (ii) some residues in the interface affected the activity of the protein *in vitro* (demonstrated through mutagenesis studies); and (iii) the structure shares structural homology with the dimeric form of outer-membrane phospholipase A (although PagL is a monomer). Modelling of the substrate lipid A on to the active site by data-driven docking reveals that the 3-*O*-acyl chain is accommodated in a hydrophobic groove perpendicular to the membrane plane. In addition, an aspartate makes a hydrogen bond with the hydroxyl group of the 3-*O*-acyl chain, probably providing specificity of PagL towards lipid A (Fig. 15.2a).

Another illustration of data-driven docking with HADDOCK is provided by the outer-membrane protein LpxR from *Salmonella typhimurium*, a Gram-negative bacterium. LpxR is a lipid A-modifying
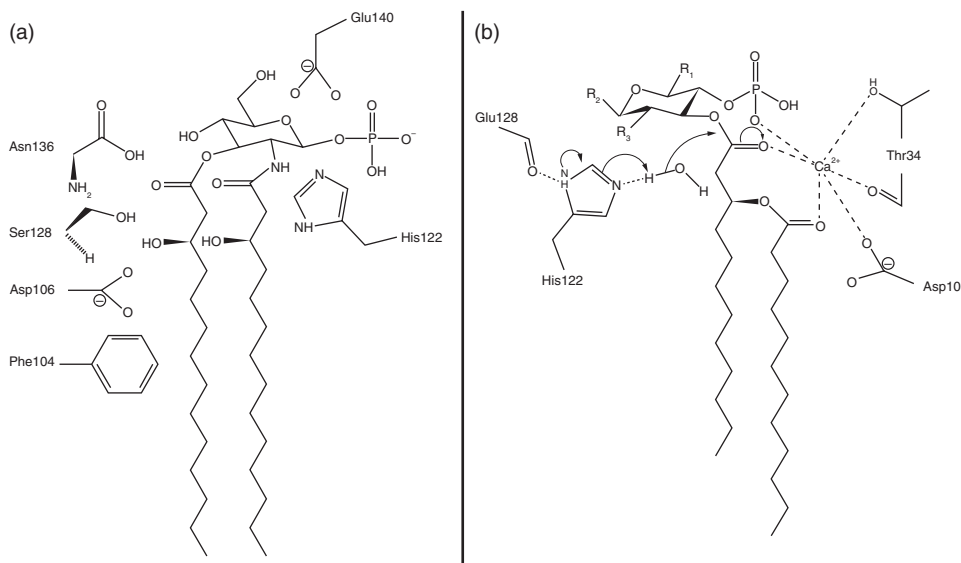
Fig. 15.2. Two-dimensional representations of docking results for lipid X molecules in outer-membrane proteins. (a) Critical interactions of lipid X and residues in the hydrophobic cleft of PagL. (b) Catalysis of lipid A by LpxR as deciphered by data-driven docking (see text).

enzyme (Rutten *et al.*, 2009) that removes the 3′-acyloxyacyl moiety of the lipid A portion of lipopolysaccharide, utilizing Ca^{2+} as a cofactor. In order to decipher its catalytic mechanism, the crystal structure of the apo-enzyme of the 32 kDa *S. typhimurium* LpxR was used as the receptor molecule. By having experimental data about residues located in the active site through mutagenesis experiments and knowing that the structure shares structural homology with phospholipase A2 (of which the catalytic mechanism is known), insight into the catalytic mechanism of LpxR could be provided. Data-driven docking was used to model the catalytic mechanism by docking lipid A to the active site of the protein, providing structural details about the recognition mechanism. Based on the derived models, the catalytic mechanism of the enzyme was established: briefly, Ca^{2+} forms the oxyanion hole and a histidine activates a water molecule (or a cascade of two water molecules) that subsequently attacks the carbonyl oxygen of the scissile bond (Fig. 15.2b).

Such detailed results for catalytic mechanisms derived for both outer-membrane proteins are unique in a sense that data-driven docking was successful in modelling the flexibility of the lipid A, a complex substrate consisting of a glucosamine (carbohydrate/sugar) unit with attached acyl chains ('fatty acids'), and one phosphate group on each carbohydrate. The resulting structural models allowed rationalization of the catalytic mechanism of these transmembrane enzymes.

15.3.2 Targeting the bacterial Achilles' heel, lipid II, using data-driven docking

The bacterial cell wall is composed of a polymerized peptidoglycan matrix that resists the high osmotic pressure of the cytoplasm, shielding the bacterium from stress. Its vital role for bacteria is also reflected by its very high conservation throughout evolution and therefore it is a prominent target for many antibiotics (Breukink and de Kruijff, 2006). The building block of the peptidoglycan matrix is the monomeric peptidoglycan unit. The latter consists of two amino sugars (*N*-acetylglycosamine and *N*-acetylmuramic acid) and a pentapeptide (commonly L-Ala-D-γ-Glu-L-Lys-D-Ala-D-Ala) attached to the carboxyl group of *N*-acetylmuramic acid. In the cellular cytosol, the membrane-anchoring

carrier undecaprenyl phosphate assembles the peptidoglycan unit parts, yielding lipid II (Fig. 15.3a). Lipid II is thereafter transported to the extracellular environment for polymerization of the peptidoglycan moiety. Many antimicrobial peptides target lipid II because of its essential role in cell-wall biosynthesis. One of the most potent inhibitors of lipid II is nisin, which interacts with lipid II through a sequence of events. First, nisin binds to the lipid II-containing membrane and forms a complex with lipid II by targeting its pyrophosphate group. It then assembles into a pore, thereby inducing leakage of cytosolic contents.

HADDOCK was used as a structure-determination program to determine the solution structure of the complex of nisin and lipid II (PDB ID: 1UZT). A wealth of distance and dihedral restraints (e.g. 619 NOEs in total) was introduced to derive the three-dimensional structure of the complex of nisin and lipid II. The structure revealed a novel lipid II-binding motif (Fig. 15.3b) in which the pyrophosphate moiety of lipid II is primarily coordinated by the N-terminal backbone amides of nisin via intermolecular hydrogen bonds and strong van der Waals interactions. Side-chain interactions thus only play a minor role in the interaction of nisin with lipid II (Hsu *et al.*, 2004), which makes it less susceptible to mutations in the peptidic region of lipid II. This cage structure offers a template for structure-based design of novel antibiotics targeting the cell-wall biosynthesis of bacteria.
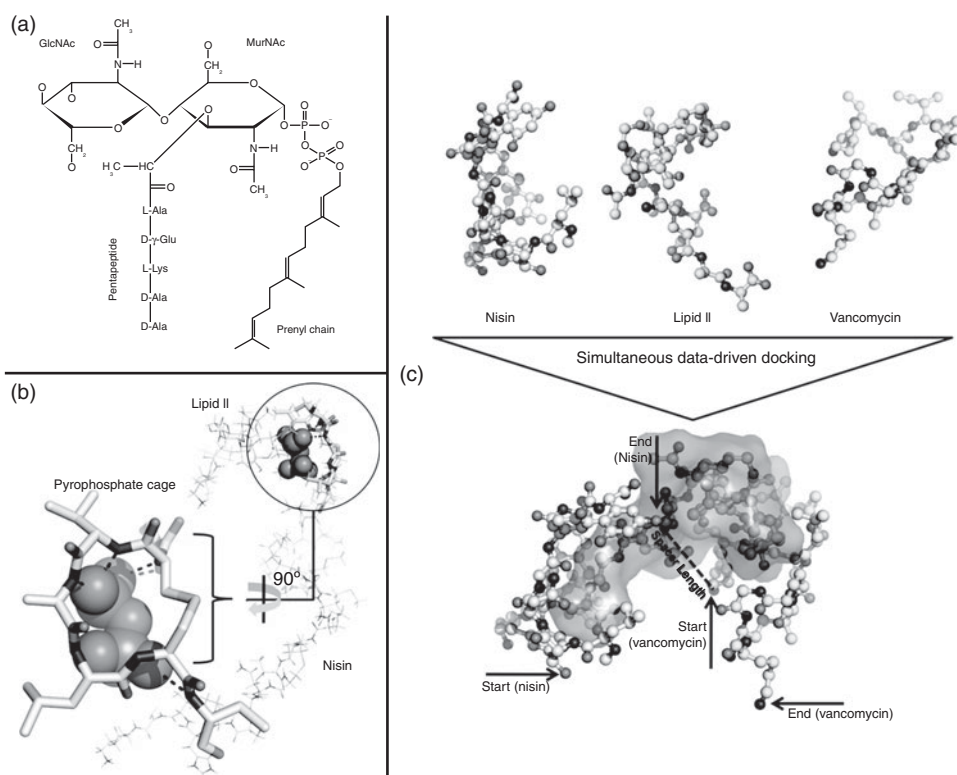
Fig. 15.3. (a) Chemical structure of the lipid II molecule. (b) Lipid II–nisin complex and the pyrophosphate cage as deciphered by HADDOCK (see text). (c) Data-driven docking of three molecules in order to derive the hybrid nisin–vancomysin molecule with click chemistry. HADDOCK was used to predict complexes and calculate spacer lengths between nisin and vancomycin (ball-and-stick representations), using lipid II as scaffold (ball-and-stick and surface representation).

Based on the structural findings for the nisin–lipid II complex, and the information for the interaction of vancomycin with the tripeptide part (Lys-D-Ala-D-Ala) of lipid II (Barna and Williams, 1984), three-body data-driven docking was applied to derive a model of the complex of both inhibitors and lipid II (Arnusch et al., 2008). The model revealed that, due to the different binding modes of vancomycin and nisin, lipid II was able to bind both molecules simultaneously. The resulting models were analysed to derive distance distributions between potential linkage points between nisin (N and C termini) and vancomycin (N and C termini) (Fig. 15. 3c). Using click chemistry, nisin and vancomycin were connected with either an alkyne or an azide group, and hybrid molecules were synthesized, one of which exhibited an antimicrobial activity superior to that of the individual agents. Data-driven docking was used in the sense that, by overcoming the obstacles of the highly sophisticated structures of the antimicrobial agents, it produced models from which spacer lengths were derived for the design of novel hybrid inhibitors of lipid II by click chemistry.

In another recent application (Schneider et al., 2010), data-driven docking with HADDOCK provided a model for the interaction between a fungal defensin (plectasin) and lipid II. This defensin can target the bacterial cell-wall precursor lipid II in a similar way to the vancomycin–nisin hybrid molecule that was designed with the aid of data-driven docking. CSP data obtained from the titration of lipid II with defensin were used to define AIRs to drive the modelling process. The model reveals that the 40 amino acid amphipathic defensin molecule binds to the solvent-exposed part of lipid II, and in particular to the pyrophosphate group. Plectasin seems thus to bind lipid II in a similar way to nisin.

physiological balance of the bile salts and the bile acids in the liver cytosol. Therefore, in the light of its biological function, it can be a suitable target for SBDD. In order to determine the ternary complex of BABP with two bile salt molecules, a hybrid computational/experimental approach was followed (Tomaselli et al., 2007). To drive the modelling process, MS and NMR data for the protein were obtained and were translated into distance restraints in HADDOCK. In this venture, a wealth of experimental information was available, including CSP, NOE and ^{15}N relaxation experiment data from NMR and limited proteolysis data from MS. During these calculations, three-body docking was performed, meaning that the two identical bile acid molecules plus the protein were docked simultaneously, making use of HADDOCK's ability to deal with multicomponent systems. A larger number of models were generated at every docking step compared with the default settings to deal with the complexity of the simultaneous three-body docking and allow a more thorough sampling of the interaction space. The resulting models revealed that residues involved in binding are mainly located in two loops at the C terminus of the protein; their orientation plays a major role in binding of the small molecules to the protein. It was also observed that polar residues pointing towards the protein interior are involved in motion communication, highlighting their prominent role in ligand interactions (Tomaselli et al., 2007).

This work emphasizes that, in such a highly demanding system where one protein can interact with two ligands simultaneously, HADDOCK can model the ternary complex, irrespective of the degree of flexibility and complexity of the system, provided enough experimental data can be used to drive the modelling process.

15.3.3 Interactions between bile acid-binding protein (BABP) and bile acids decrypted by NMR, MS and computational modelling

Although not a potential target for antimicrobial agents, BABP is a key element in cholesterol homeostasis as it controls the

15.3.4 Scoring in HADDOCK as an estimate of binding energies of peptide inhibitors for the treatment of HIV-1 viral infections

HIV-1 attachment to CD4$^+$ target T cells and subsequent fusion of viral and cellular

membranes resulting in release of the viral core into the cell is accomplished by the HIV-1 envelope glycoprotein complex (Env), a class I viral fusion protein. HIV-1 Env is a trimer, with each monomer consisting of two subunits, proteins gp120 and gp41. Whereas gp120 is responsible for adhesion and (partially) fusion, gp41 induces fusion of the viral envelope with the plasma membrane, thereby initiating infection. Each gp41 molecule contains an N-terminal leucine/isoleucine heptad repeat (HR) segment that has been crystallographically shown to form a central triple-stranded α-helical coiled-coil core (Weissenhorn *et al.*, 1997). Peptides based on the second heptad repeat (HR2) of viral class I fusion proteins are effective inhibitors of virus entry. For example, a fusion inhibitor has been approved for treatment of HIV infections (T20, or enfuvirtide).

For this kind of system, only the water refinement part of the HADDOCK protocol was used to extract the energetics of protein–peptide complexes of modelled wild-type and mutant gp41 proteins, in complex with three different inhibitors (including enfuvirtide).

The derived modelled complexes were constructed based on homology with existing crystallographic structures (Caffrey *et al.*, 1998), in light of mutational experiments that were mapped on to the models. The intermolecular energies of the complexes were obtained and compared with experimental knowledge about the resistance of mutant gp41 proteins to the inhibitory peptides. This allowed the proposition of four different mechanisms of resistance to fusion inhibitors. Although scoring should not be interpreted in terms of affinity (Kastritis and Bonvin, 2010), the modelling data suggested that the presence of exposed charges on the peptide at the drug–target interface, unless involved in an intramolecular salt bridge, is not desirable. Exposed charges may provide the virus with an easy possibility to generate the most powerful mechanism of resistance – electrostatic repulsion. Such findings may guide the design of novel fusion inhibitors targeting viruses with class I fusion proteins.

15.4 Conclusions

As the antimicrobial drug discovery community moves towards more complicated systems, in terms of both the ligands and the targets, relevant computational methods are needed to tackle the challenges in modelling their three-dimensional structures and interaction mode. The trend towards increased molecular weight of the designed agents goes together with an increase in degrees of freedom and underlying flexibility of the biomolecules. In addition, molecules are becoming more and more complicated, being built of different chemical groups from different classes of chemical molecules such as, for example, those targeting lipid II biosynthesis. On the receptor side, the targets are getting also more complex. For example, protein–protein interactions and their interfaces are becoming relevant for the development of novel therapeutics. The plasticity and flexibility of some interfaces can be a challenge, even for the most sophisticated docking programs. Recognition events mediated by local folding or unfolding of biomolecules, large loop rearrangements and binding-affinity prediction of the modelled interactions still represent three major bottlenecks that docking programs have to face in the years to come.

Despite its limitations, data-driven docking offers a good solution to tackle some of these challenges, as highlighted by the various applications described in the previous sections. Provided enough experimental data are available, HADDOCK can work as a structure-determination program, able to generate three-dimensional structures of biomolecular complexes, acting as a catalyst for deciphering biomolecular interactions. This is illustrated by the fact that, as of February 2012, 85 macromolecular complexes have been deposited into the PDB for which HADDOCK has been used as a structure-determination package.

Manuals and Tutorials on Data-driven Docking

Further information on the use of HADDOCK can be found as follows:

1. HADDOCK: http://www.wenmr.eu/wenmr/
tutorials/nmr-tutorials/haddock. In this
website, several tutorials related to data-
driven docking can be found:

- A case study: preparing input files for a
  manual HADDOCK run
- Generating the necessary restraint files
  for running HADDOCK manually
- How to prepare PDB files for running
  HADDOCK manually
- HADDOCK web server tutorial

2. A demo web form for the easy interface
with pre-loaded parameters: http://haddock.
science.uu.nl/enmr/services/HADDOCK/
haddockserver-demo.html.

3. Tutorial movie on data-driven docking:
http://haddock.science.uu.nl/Files/e2a-
hpr-demo.swf.

Acknowledgements

References

Arkin, M.R. and Wells, J.A. (2004) Small-molecule
inhibitors of protein–protein interactions: pro-
gressing towards the dream. *Nature Reviews
Drug Discovery* 3, 301–317.

Arnusch, C.J., Bonvin, A.M., Verel, A.M., Jansen, W.T.,
Liskamp, R.M., de Kruijff, B., Pieters, R.J. and
Breukink, E. (2008) The vancomycin–nisin(1–12)
hybrid restores activity against vancomycin resist-
ant enterococci. *Biochemistry* 47, 12661–12663.

Barna, J.C. and Williams, D.H. (1984) The structure
and mode of action of glycopeptide antibiotics
of the vancomycin group. *Annual Reviews in
Microbiology* 38, 339–357.

Becattini, B. and Pellecchia, M. (2006) SAR by
ILOEs: an NMR-based approach to reverse
chemical genetics. *Chemistry* 12, 2658–2662.

Betz, M., Saxena, K. and Schwalbe, H. (2006)
Biomolecular NMR: a chaperone to drug dis-
covery. *Current Opinion in Chemical Biology* 10,
219–225.

Blundell, T.L., Jhoti, H. and Abell, C. (2002) High-
throughput crystallography for lead discovery in
drug design. *Nature Reviews Drug Discovery* 1,
45–54.

Bourgeas, R., Basse, M.J., Morelli, X. and Roche, P.
(2010) Atomic analysis of protein–protein inter-
faces with known inhibitors: the 2P2I database.
PLoS One 5, e9598.

Breukink, E. and de Kruijff, B. (2006) Lipid II as
a target for antibiotics. *Nature Reviews Drug
Discovery* 5, 321–332.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano,
W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang,
J.S., Kuszewski, J., Nilges, M., Pannu, N.S.,
Read, R.J., Rice, L.M., Simonson, T. and Warren,
G.L. (1998) Crystallography and NMR system:
a new software suite for macromolecular struc-
ture determination. *Acta Crystallographica D:
Biological Crystallography* 54, 905–921.

Caffrey, M., Cai, M., Kaufman, J., Stahl, S.J.,
Wingfield, P.T., Covell, D.G., Gronenborn, A.M.
and Clore, G.M. (1998) Three-dimensional solu-
tion structure of the 44 kDa ectodomain of SIV
gp41. *EMBO Journal* 17, 4572–4584.

Cowan, M.M. (1999) Plant products as antimicro-
bial agents. *Clinical Microbiology Reviews* 12,
564–582.

de Vries, S.J. and Bonvin, A.M. (2008) How proteins
get in touch: interface prediction in the study of
biomolecular complexes. *Current Protein and
Peptide Science* 9, 394–406.

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

de Vries, S.J., Melquiond, A.S., Kastritis, P.L.,
Karaca, E., Bordogna, A., van Dijk, M.,
Rodrigues, J.P. and Bonvin, A.M. (2010a)
Strengths and weaknesses of data-driven
docking in critical assessment of prediction of
interactions. *Proteins: Structure, Function and
Bioinformatics* 78, 3242–3249.

de Vries, S.J., van Dijk, M. and Bonvin, A.M. (2010b)
The HADDOCK web server for data-driven bio-
molecular docking. *Nature Protocols* 5, 883–897.

Dominguez, C., Boelens, R. and Bonvin, A.M.
(2003) HADDOCK: a protein–protein docking
approach based on biochemical or biophysical
information. *Journal of the American Chemical
Society* 125, 1731–1737.

Dorsey, B.D., Levin, R.B., McDaniel, S.L., Vacca,
J.P., Guare, J.P., Darke, P.L., Zugay, J.A., Emini,

E.A., Schleif, W.A., Quintero, J.C., Quintero, J.C., Linn, J.H., Chen, I.W., Holloway, M.K., Fitzgerald, P.M.D., Axel, M.G., Ostovic, D., Anderson, P.S. and Huff, J.R. (1994) L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor. *Journal of Medicinal Chemistry* 37, 3443–3451.

Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M. (1990) Design, activity, and 2.8 Å crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science* 249, 527–533.

Fiamegos, Y.C., Kastritis, P.L., Exarchou, V., Han, H., Bonvin, A.M.J. J., Vervoort, J., Lewis, K., Hamblin, M.R. and Tegos, G.P. (2011) Antimicrobial and efflux pump inhibitory activity of caffeoylquinic acids from *Artemisia absinthium* against Gram-positive pathogenic bacteria. *PLoS One*, 6, e18127.

Fuentes, G., Dastidar, S.H., Madhumalar, A. and Verma, C.S. (2011) Role of protein flexibility in the discovery of new drugs. *Drug Development Research* 72, 26–35.

Grunberg, R., Leckner, J. and Nilges, M. (2004) Complementarity of structure ensembles in protein–protein binding. *Structure* 12, 2125–2136.

Hajduk, P.J. and Greer, J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug Discovery* 6, 211–219.

Hann, M.M., Leach, A.R. and Harper, G. (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of Chemical Information in Computer Sciences* 41, 856–864.

Hsu, S.T., Breukink, E., Tischenko, E., Lutters, M.A., de Kruijff, B., Kaptein, R., Bonvin, A.M. and van Nuland, N.A. (2004) The nisin–lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. *Nature Structure and Molecular Biology* 11, 963–967.

Jahnke, W. (2007) Perspectives of biomolecular NMR in drug discovery: the blessing and curse of versatility. *Journal of Biomolecular NMR* 39, 87–90.

Jenssen, H., Hamill, P. and Hancock, R.E. (2006) Peptide antimicrobial agents. *Clinical Microbiology Reviews* 19, 491–511.

Jung, S., Dingley, A.J., Augustin, R., Anton-Erxleben, F., Stanisak, M., Gelhaus, C., Gutsmann, T., Hammer, M.U., Podschun, R., Bonvin, A.M., Leippe, M., Bosch, T.C. and Grotzinger, J. (2009) Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan Hydra. *Journal of Biological Chemistry* 284, 1896–1905.

Karaca, E. and Bonvin, A.M. (2011) A multi-domain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure* 19, 555–565.

Karaca, E., Melquiond, A.S., de Vries, S.J., Kastritis, P.L. and Bonvin, A.M. (2011) Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Molecular and Cellular Proteomics* 9, 1784–1794.

Kastritis, P.L. and Bonvin, A.M. (2010) Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *Journal of Proteome Research* 9, 2216–2225.

Kastritis, P.L., van Dijk, A.-J. and Bonvin, A.M.J.J. (2012) Explicit treatment of water molecules in data-driven docking: the Solvated HADdocking approach. *Methods in Molecular Biology* 819, 5–374.

Kolb, H.C. and Sharpless, K.B. (2003) The growing impact of click chemistry on drug discovery. *Drug Discovery Today* 8, 1128–1137.

Krzeminski, M., Singh, T., Andre, S., Lensch, M., Wu, A.M., Bonvin, A.M. and Gabius, H.J. (2010) Human galectin-3 (Mac-2 antigen): defining molecular switches of affinity to natural glycoproteins, structural and dynamic aspects of glycan binding by flexible ligand docking and putative regulatory sequences in the proximal promoter region. *Biochimica et Biophysica Acta* 1810, 150–161.

Lensink, M.F. and Wodak, S.J. (2010) Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins: Structure, Function and Bioinformatics* 78, 3085–3095.

McCauley, J. (1999) Relenza. *Current Biology* 9, R796.

Meagher, K.L. and Carlson, H.A. (2004) Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. *Journal of the American Chemical Society* 126, 13276–13281.

Melquiond, A.S.J. and Bonvin, A.M.J.J. (2010) Data-driven docking: using external information to spark the biomolecular rendez-vous. In: Zacharias, M. (ed.) *Protein–Protein Complexes: Analysis, Modelling and Drug Design*. Imperial College Press, Munich, Germany, pp. 183–209.

Murray, C.W. and Blundell, T.L. (2010) Structural biology in fragment-based drug design. *Current Opinion in Structural Biology* 20, 497–507.

Patel, S. and Player, M.R. (2008) Small-molecule inhibitors of the p53–HDM2 interaction for the treatment of cancer. *Expert Opinion on Investigational Drugs* 17, 1865–1882.

Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B.,

Galpin, S.A., Handa, B.K., Kay, J., Krohn, A., Lambert, R., Merrett, J., Mills, J, Parks, K., Redshaw, S., Taylor, D., Thomas, G. and Machin, P. (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science* 248, 358–361.

Rutten, L., Geurtsen, J., Lambert, W., Smolenaers, J.J., Bonvin, A.M., de Haan, A., van der Ley, P., Egmond, M.R., Gros, P. and Tommassen, J. (2006) Crystal structure and catalytic mechanism of the LPS 3-O-deacylase PagL from *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences USA* 103, 7071–7076.

Rutten, L., Mannie, J.P., Stead, C.M., Raetz, C.R., Reynolds, C.M., Bonvin, A.M., Tommassen, J.P., Egmond, M.R., Trent, M.S. and Gros, P. (2009) Active-site architecture and catalytic mechanism of the lipid A deacylase LpxR of *Salmonella typhimurium*. *Proceedings of the National Academy of Sciences USA* 106, 1960–1964.

Sanchez-Pedregal, V.M., Reese, M., Meiler, J., Blommers, M.J., Griesinger, C. and Carlomagno, T. (2005) The INPHARMA method: protein-mediated interligand NOEs for pharmacophore mapping. *Angewandte Chemie International Edition* 44, 4172–4175.

Sawaya, M.R. and Kraut, J. (1997) Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* 36, 586–603.

Schmitz, C., Melquiond, A.S.J., de Vries, S.J., Karaca, E., van Dijk, M., Kastritis, P.L. and Bonvin, A.M.J.J. (2012) Chapter 9.2. HADDOCK. In: Bertini, I., McGreevy, K.S. and Parigi, G. (eds) *NMR of Biomolecules: Towards Mechanistic Systems Biology*. Wiley, Chichester, UK (in press).

Schneider, T., Kruse, T., Wimmer, R., Wiedemann, I., Sass, V., Pag, U., Jansen, A., Nielsen, A.K., Mygind, P.H., Raventos, D.S., Neve, S., Ravn, B., Bonvin, A.M., De Maria, L., Andersen, A.S., Gammelgaard, L.K., Sahl, H.G. and Kristensen, H.H. (2010) Plectasin, a fungal defensin, targets the bacterial cell wall precursor Lipid II. *Science* 328, 1168–1172.

Schuttelkopf, A.W. and van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallographica D Biological Crystallography* 60, 1355–1363.

Simmons, K.J., Chopra, I. and Fishwick, C.W. (2010) Structure-based discovery of antibacterial drugs. *Nature Reviews Microbiology* 8, 501–510.

Song, J., Zhao, K.Q., Newman, C.L., Vinarov, D.A. and Markley, J.L. (2007) Solution structure of human sorting nexin 22. *Protein Science* 16, 807–814.

Stratton, M.S. and Alberts, D.S. (2002) Current application of selective COX-2 inhibitors in cancer prevention and treatment. *Oncology (Williston Park)* 16 (Suppl. 4), 37–51.

Tomaselli, S., Ragona, L., Zetta, L., Assfalg, M., Ferranti, P., Longhi, R., Bonvin, A.M. and Molinari, H. (2007) NMR-based modeling and binding studies of a ternary complex between chicken liver bile acid binding protein and bile acids. *Proteins: Structure, Function and Bioinformatics* 69, 177–191.

Toogood, P.L. (2002) Inhibition of protein–protein association by small molecules: approaches and progress. *Journal of Medicinal Chemistry* 45, 1543–1558.

van Dijk, A.D. and Bonvin, A.M. (2006) Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* 22, 2340–2347.

van Dijk, A.D., Boelens, R. and Bonvin, A.M. (2005) Data-driven docking for the study of biomolecular complexes. *FEBS Journal* 272, 293–312.

Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics* 8, 52–56.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J. and Wiley, D.C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387, 426–430.

Wu, A.M., Singh, T., Liu, J.H., Krzeminski, M., Russwurm, R., Siebert, H.C., Bonvin, A.M., Andre, S. and Gabius, H.J. (2007) Activity–structure correlations in divergent lectin evolution: fine specificity of chicken galectin CG-14 and computational analysis of flexible ligand docking for CG-14 and the closely related CG-16. *Glycobiology* 17, 165–184.