

REVIEW ARTICLE

Data-driven docking for the study of biomolecular complexes

Aalt D. J. van Dijk, Rolf Boelens and Alexandre M. J. J. Bonvin

Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, the Netherlands

Keywords

biomolecular complexes; docking; interface mapping

Correspondence

A. M. J. J. Bonvin, Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, 3584CH, Utrecht, the Netherlands
Fax: +31 (0) 30 2537623
Tel: +31 (0) 30 2532652
E-mail: a.m.j.j.bonvin@chem.uu.nl
Website: <http://www.nmr.chem.uu.nl>

(Received 1 October 2004, revised 5 November 2004, accepted 10 November 2004)

doi:10.1111/j.1742-4658.2004.04473.x

With the amount of genetic information available, a lot of attention has focused on systems biology, in particular biomolecular interactions. Considering the huge number of such interactions, and their often weak and transient nature, conventional experimental methods such as X-ray crystallography and NMR spectroscopy are not sufficient to gain structural insight into these. A wealth of biochemical and/or biophysical data can, however, readily be obtained for biomolecular complexes. Combining these data with docking (the process of modeling the 3D structure of a complex from its known constituents) should provide valuable structural information and complement the classical structural methods. In this review we discuss and illustrate the various sources of data that can be used to map interactions and their combination with docking methods to generate structural models of the complexes. Finally a perspective on the future of this kind of approach is given.

Introduction

With the available amount of genetic information, a lot of attention is focused on systems biology. Here a central question is: how do the various biomolecular units work together to fulfil their tasks? To answer this question, structural information on complexes is needed. Biochemical and biophysical experiments are widely used to gain insight into biomolecular interactions. The information generated in this way can in principle be used to model the structure of the complex under study. Taking the step from data to modeling (docking) is, however, not common practice. Docking approaches allow models of a biomolecular complex to be generated using as starting information the known structure of its constituents. Combining experimental data with docking makes sense considering that the

number of single proteins, domains thereof, or other biomolecules whose 3D structures have been solved is much larger than the number of solved structures of complexes and is steadily increasing as a result of the worldwide structural genomics initiatives. The advantages of docking approaches over conventional structural techniques are the speed and the possibility of studying complexes that could only otherwise be studied with considerable effort (or not at all). One particular class of complexes for which this is the case are weak or transient, short-lived complexes; this is all the more interesting as these are often of the utmost biological importance. Other examples are the biologically highly relevant complexes of membrane or membrane-associated proteins, which are also notoriously difficult to study by NMR spectroscopy or X-ray crystallography.

Abbreviations

AIR, ambiguous interaction restraint; CAPRI, critical assessment of predicted interactions; CSP, chemical shift perturbation; HADDOCK, high ambiguity driven docking; HSQC, heteronuclear single quantum coherence; RDC, residual dipolar coupling; SAXS, small angle X-ray scattering.

Conventional crystallographic and NMR structural biology techniques have proven their value and will continue to do so. There are, however, problems associated with these techniques that are not likely to be completely overcome, especially when dealing with complexes. For crystallography, the main bottleneck is the crystallization, which can be a daunting task. For NMR, large complexes cause severe line broadening, which, at present, sets the upper limit for NMR to molecular sizes below 100 kDa. Moreover, to solve a structure by NMR in a conventional way, complete chemical shifts assignment and collection of structural restraints such as NOEs are challenging tasks, especially for large systems such as complexes.

In this review, we wish to highlight the use of biochemical and biophysical data in docking approaches not only because of the general interest in docking as explained above, but also because it is still common practice to experimentally map interfaces without taking the next step of generating a structural model of the complex. We review only part of the docking field, namely approaches that rely on the use of additional biochemical and/or biophysical data. Generally, docking approaches that do not use any kind of experimental data have difficulty in generating consistently reliable structures of complexes. Nevertheless, clear progress has been achieved in the field of ‘*ab-initio* docking’, as reviewed in [1–4], and illustrated by the critical assessment of predicted interactions (CAPRI) experiment [5], a ‘blind’ docking competition in which participants have a limited time to predict the structure of a complex given only the structures of the constituents. Our discussion will be limited to biomolecular complexes, omitting protein–small ligand complexes; however, much of what is presented here will also be valid for that class of complexes. For a review on ‘guided docking’ for studying protein–ligand complexes, see reference [6].

The review is organized as follows. We will first discuss the various kinds of biochemical and biophysical data that can be combined with docking. For each of these, examples will be given, and their strengths and weaknesses for use in docking will be discussed. We will then describe the basics of current docking methodologies and highlight our newly developed data-driven docking method HADDOCK [7]. We will end with conclusions and give a broader perspective on what could be the future of data-supported docking.

Sources of experimental data to define interfaces

Data from biochemical and/or biophysical experiments that provide information on residues located at the interface of a complex are potential sources to be used

in docking. Critical issues are the level of detail that can be obtained (e.g. is the information residue-specific or not?) and the reliability of the data. Here we discuss, with those issues in mind, the techniques that have been used to obtain interface information for docking. In Fig. 1 we present an overview of the most common methods. For a selected set of examples, we will also discuss how these data relate to the experimental high-resolution structure solved by conventional methods (Table 4). Other experimental methods such as small angle X-ray scattering (SAXS) or electron microscopy and tomography can also provide valuable information about the ‘shape’ and organization of biomolecular complexes. As these are rather different kinds of approaches, we will not review them here, but only briefly mention their potential in our conclusions and perspectives. A general review of structural perspectives on protein–protein interactions can be found in reference [8].

Mutagenesis

When using mutagenesis to derive information for docking, one considers as candidates only the residues that are on the surface of the partner proteins. The general idea then is that mutation of an interface residue will influence the interaction, whereas for non-interface residues the mutation will have no effect. A variety of methods can be used to find out whether complex formation is affected by mutations, such as surface plasmon resonance [9], MS, yeast two-hybrid systems [10] and phage display libraries [11]. Target residues for mutagenesis can be selected based on knowledge such as conservation (see below), but it is also possible to perform an in-depth systematic scan as in alanine scanning mutagenesis studies [12,13]. An online database with results from alanine scanning mutagenesis has been established called ASeddb (<http://www.aseddb.org>) [14]. These methods indicate which residues are in the interface, but do not give information about the contacts that are made across the interface. More detailed information can be obtained using so-called double mutant cycles [15]. Here one creates a series of mutants for both proteins. By measuring the K_d values for combinations of mutants, one can assess whether the influence of mutation X in protein A on the complex formation depends on mutation Y in protein B. If this is the case, the mutations are coupled, and one infers that the residues are close in space, i.e. that they are in contact or close proximity across the interface.

A general warning when using mutagenesis data is that it is unsound to assume that residues for which no effect is seen on mutation do not participate in

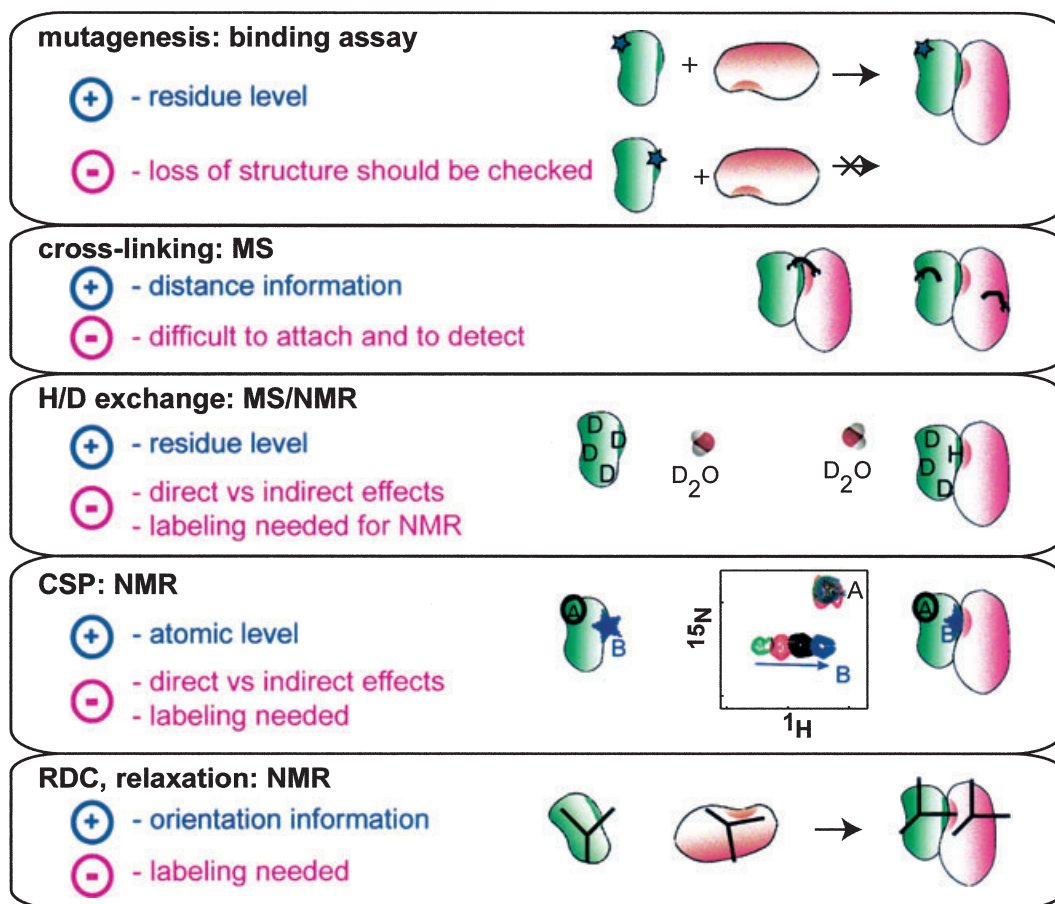


Fig. 1. Illustration of the various data sources used in combination with docking. Left: advantages (+) and disadvantages (-); right: pictorial representation of the data source: the green and red shapes represent the two components of the complex. Mutagenesis: the blue star indicates a mutated residue; cross-linking: the black line indicates a cross-link; H/D exchange: 'D' and 'H' indicate residues where exchange can and cannot take place, respectively; CSP (chemical shift perturbation): HSQC spectrum showing one peak that does not shift and one peak that shifts on complex formation (the corresponding residues are indicated on the protein shapes); RDC, relaxation: the axis system indicates the tensor which provides orientational information.

an important interaction, unless it can be demonstrated that water, or nearby side chains, do not effectively substitute for the deleted atoms [13]. Another point is that one should, in principle, always check whether the mutants do not affect the 3D structure of the free components themselves, i.e. whether or not the native structures are preserved. Mutagenesis approaches, when carried out extensively, are able to generate a fairly detailed map of the interface of a biomolecular complex. In Table 1 we give an overview of complexes for which mutagenesis data have been used in docking.

Mass spectrometry

There has been increasing interest in MS as a tool in structural biology in general, and also specifically to

obtain information about biomolecular complexes [16,17]. One approach that can be used is H/D exchange. Here the rate of exchange gives information about the accessibility of the residue in question; rate differences between free and bound forms indicate that a given residue is protected on complex formation and thus probably involved in the interaction [18,19]. Another possibility is cross-linking, where residues close in space are detected by first covalently linking two molecules by the use of a cross-linking reagent, and then subjecting the resulting material to peptide mass fingerprinting or other protein identification methods [20]. Although these methods are promising, the cross-linking reaction is problematic, and the information is often not easy to interpret. The detection of cross-linked residues is especially nontrivial. To date MS data have not often yet been combined with docking approaches (Table 2).

Table 1. Examples of complexes docked using mutagenesis data (GST, glutathione *S*-transferase; SPR, surface plasmon resonance; CSP, chemical shift perturbation). –, Data were taken from the literature without giving any experimental details.

Complex	Information used	Reference
Mutagenesis		
FAK FAT domain–paxillin-derived LD2 peptide	GST domain fusion	[89]
TF/fvIIa/fXa	Charge altering mutations	[152]
R ₁₁ α–C _α subunits of PKA	Neutron scattering, mutagenesis	[110]
SDF-1α–heparin	SPR	[153]
RCC1–Ran	SPR	[51]
Glycophorin A dimer	–	[45]
Phospholamban pentamer	–	[44,154,155]
Staphylokinase–microplasmin	Phage display	[156]
Gα–Gβγ-receptor	G-protein activation assay	[157]
30S ribosomal subunit–colicin E3	Immunoblotting	[70,71]
EmrE dimer	Cysteine mutagenesis, cross-linking	[78]
Hsc70–auxilin	Rescue-mutant pair, CSP	[158]
Kv1.3 K ⁺ channel αIIb – six different scorpion toxins	Comparison of electrostatic energy with binding affinity	[63]
Integrin αIIb TM domain homodimer	CAT-ELISA	[47]
C1q–C-reactive protein/IgG	–	[49]
Antibody fragment–α bungarotoxin	CDR on antibody; epitope mapping	[159]
Malonyl-CoA–COT/CPT	Enzyme activity assay, immunoblotting	[160]
gp120–CD4	–	[7]
Protein–DNA complexes of 434 cro and lac headpiece	Ethylation interference	[34]
LexA DBD–DNA	Ethylation interference	[72]
LexA–DNA	Cross-linking	[161]
Repressor–protein–DNA	DNA footprinting	[162]
Fis–DNA	Chemical interference, nuclease DNA cleavage site	[163]
EnvZ dimer	Cysteine substitutions and disulfide cross-linking detection	[164]
Subunit c oligomer of H ⁺ -transporting ATP synthase	Cysteine substitutions and disulfide cross-linking detection	[165]
Yeast cofactor A–β-tubulin	Two-hybrid assay	[166]
FOG-ZF3 _{KRA} –TACC3	Two-hybrid assay; NMR CSP	[90]
Double mutant cycles		
BgK–Kv1.1	Electrophysiological experiments, dose–response curve	[74]
Agitoxin– <i>shaker</i> K ⁺ channel	–	[75]
IFN-α2–ifnar2	Reflectometric interference spectroscopy	[77]
α-Cobratoxin–α7 receptor	Binding competition	[76]

Table 2. Examples of complexes docked using MS data.

Complex	Information used	Reference
Calmodulin–melittin	Cross-linking	[85]
Aminoacylase-1 dimer	Proteolysis, cross-linking	[111]
PKA–C and R subunit	H/D exchange	[50]
C1r (γ-B) ₂	Cross-linking	[167]
IL-6 homodimer	Cross-linking	[112]

NMR

Conventional NMR methods have been used for more than a decade to study biomolecular complexes. In the classical approach, one first has to perform a resonance assignment that is as complete as possible, and then collect structural restraints such as NOEs, which can be detected between protons that are close in space ($< 5 \text{ \AA}$), and residual dipolar couplings that provide

orientational information. Using such restraints, one can accurately define the structure of a biomolecule or a biomolecular complex. In addition to its conventional use in structure determination, NMR is very well suited to map interfaces of biomolecular complexes with so-called chemical shift perturbation (CSP) experiments [21]. Here, easily obtainable heteronuclear single quantum coherence (HSQC) spectra of one (¹⁵N-labeled) partner in the complex are recorded in the absence and presence of increasing amounts of the partner protein ('titration experiments'). Changes in chemical shifts of one molecule on addition of a second molecule allow assessment of which residues of the labeled molecule are perturbed by the formation of the complex. One then repeats this procedure with the second molecule labeled. Under the assumption that the perturbed residues correspond to the interacting residues, a detailed map of the interface is obtained.

Two other NMR techniques that are able to give similar information are H/D exchange and cross-saturation or saturation transfer [22]. As in MS, NMR can also easily be used to perform H/D exchange experiments; again, differences in exchange rates when comparing uncomplexed and complexed forms point to protected residues that are assumed to be at the interface. In cross-saturation experiments, the observed protein is perdeuterated and ^{15}N -labeled, with its amide deuterons exchanged back to protons, while the other 'donating' partner protein is unlabeled. Saturation of the unlabeled protein leads by cross-relaxation mechanisms to signal attenuation (again typically monitored by ^{15}N -HSQC spectra) of those residues in the labeled protein that are in close proximity. The labeling scheme can be reversed to map the other interface. Deuteration is a requisite here. Cross-saturation experiments are believed to give a more reliable picture of the interface than CSP data, which can suffer from 'false positives' because of conformational changes.

Other relatively easily obtainable NMR parameters are residual dipolar couplings (RDCs) [23]. These provide information about the orientation of the components with respect to each other, and can be used in addition to CSP data in docking approaches. Comparable information can be extracted from relaxation experiments in the case of diffusion anisotropy [24].

A NMR parameter that can also be useful is the pseudocontact shift. It results from residual electron-nuclei dipolar interactions in molecules [21]. The use of paramagnetic tags attached to a protein can induce this phenomenon [25,26]. As pseudocontact shifts contain long-range information, they can be very useful in docking approaches. It is also possible to use paramagnetic ions as probes, as they induce broadening of the NMR signals for the residues they contact. In a complex, the interface residues will be protected from such effects, allowing a reliable detection of the interface [27]. An overview of complexes for which NMR data have been used in docking approaches is given in Table 3.

Reliability issues

It should be clear that there is a wealth of experimental data, not all of them having been discussed here, that can be used to define interface residues. The question of the reliability of this information is of course very important. In Table 4 we give an overview of some complexes for which the experimental data have been compared explicitly with the (at that time available) corresponding 3D structures. In Fig. 2, as an example, experimental data for the antibody D1.3-antibody E5.2 complex is mapped on to the surfaces

Table 3. Examples of complexes docked using NMR data (CSP, chemical shift perturbation; PC, pseudocontact shifts; SAT, saturation transfer).

Complex	Information used	Reference
Protein-protein		
Cyt c-cyt <i>f</i>	CSP	[56]
Cyt c-cyt <i>c</i> peroxidase	CSP	[54]
Plastocyanin-cyt <i>f</i>	PC, CSP	[80,81]
Myoglobin-cyt <i>b5</i>	CSP, ^{15}N relaxation	[57]
Ubiquitin-YUH1	CSP	[38]
Ubiquitin-hHR23A UBA1, UBA2	CSP	[93]
hHR23a (four linked domains)	CSP, RDC	[168]
Ubiquitin-p47 UBA domain	CSP	[96]
Di-ubiquitin	CSP, RDC	[169,170]
UbcH5B-CNOT4	CSP, mutagenesis	[88]
mms2-ubc13-ubiquitin-ubiquitin	CSP	[59]
EIN-HPr ^a , IIA(Glc)-HPr ^a , IIA(Mtl)-HPr ^a	CSP, RDC	[84]
Bem1 PB1-Cdc24 PB1	CSP, mutagenesis	[95]
RPA70A-Rad51N	CSP, mutagenesis	[94]
CAD-ICAD ^a	SAT, RDC	[82]
EIN-HPr ^a	CSP, RDC	[67]
EIN-HPr ^a , E2A-HPr ^a	CSP	[7]
Atx1-Ccc2 domain	CSP	[92]
HR1b-Rac1	CSP	[171]
FcεRIα-IgE Cε2	CSP	[172]
FcεRI-peptide	CSP, mutagenesis, NOE	[66]
LpxA-acyl carrier protein	CSP, RDC, mutagenesis	[91]
Protein-carbohydrates		
Tri,hexa saccharide-antibody	SAT	[173]
(Glycosylated)	SAT	[174]
PDTRP-antibody SM3		
Fibronectin (13,14)F3-heparin	CSP	[62]
Protein-nucleic acids		
NS1A(1-73)-16 bp dsRNA	CSP	[40]
UvrC CTD-junction DNA	CSP	[39]
XPA-MBD-9 bp ssDNA	CSP	[175]
Rom-RNA kissing hairpin	CSP	[41]
Pf3 ssDBP-ssDNA	CSP	[83]
CylR2-22 bp DNA	CSP	[73]

^a These complexes were also solved using the classical NOE-based approach.

of the two proteins. Although these are only a few examples, the general trend indicates that the experimental sources discussed above provide quite reliable information on interface residues. Sometimes they can result from small rearrangements and secondary effects, but as long as these 'false positives' are not too numerous, they can be dealt with in computational approaches (see below). If conformational changes are too large, however, docking approaches are probably bound to fail. It is not simple to predict a priori from the data if such effects should be expected. Sometimes,

Table 4. Comparison of experimental information defining interfaces with the experimental X-ray or NMR structures (CSP, chemical shift perturbation; DMC, double mutant cycles; SAT, saturation transfer).

Complex	Information used	Reference
Mutagenesis data		
Barnase–barstar	DMC: coupling energy decreases as distance increases	[176]
Antibody D1.3–antibody E5.2	DMC: of 13 identified, 9 in interface and 4 not in interface showing significant coupling, but lower than the contacting residues	[177]
Cyt <i>c</i> –peroxidase	Mutations: sites coincide with X-ray defined sites; DMC: couplings for residues that are more than 10 Å apart, concluded to be due to small rearrangements	[178]
Cyt <i>c</i> 2–RC	DMC: coupling approximately inversely proportional to distances	[179]
MS data		
DnaA domain 4–DnaA box	Cross-linking data correctly locate the interaction site to a six residue peptide fragment identified previously by X-ray/NMR	[180]
Ribosome	Comparison of > 2500 experimental distance restraints (cross-linking, footprinting and cleavage data) with X-ray structure showing good agreement	[144]
NMR data		
Lysozyme–antibody	H/D: of 15 perturbed: 5 on epitope, 5 at edge, 5 far away	[181]
OMTKY3–Ctr	CSP fully consistent with X-ray	[182]
rNTF2–FxFG-containing Nsp1–P30	High affinity X-ray site seen by NMR; NMR also finds low affinity site → NMR data better able to identify weak interactions	[183]
Zf1–3 (TFIIIA)–15 bp DNA	CSP data do not correspond exactly to the interface, but arise from a number of effects	[184]
CAD–ICAD	NOE and SAT defined interface is quite consistent with X-ray; CSP defined interface is a bit different	[82]
Nova1–RNA	Cross-saturation defined residues match closely the X-ray interface; CSP data define the same residues and a few additional ones	[185]
RNAse E S1 homodimer	CSP used to assess validity of crystallography dimer; data match the contacting residues seen in the crystal	[186]

clustering of predicted interface residues on the surface can give a good indication that the mapped interface is very likely to be the correct one.

Computational docking approaches using experimental data

In the docking literature one often finds the distinction between ‘bound’ and ‘unbound’ docking: the former refers to docking using the structures of the single proteins as they are present in the complex, and the latter to docking using the structures of the free proteins. As only the latter is of biological relevance, here ‘docking’ will refer to ‘unbound docking’ (although in some cases a method is, as a first, easier step, tested in bound docking).

As defined in the introduction, docking methods generate a model of a complex based on the known 3D structures of its free components. To do this in a computer, two things are needed: a way to generate structures of the complex, i.e. a sampling method, and a way to decide which of the generated structures are ‘good’, i.e. a scoring method. The output typically consists of a large number of solutions, some of which get a high ranking and are accordingly considered to correspond to the ‘real’ structure, whereas others get a lower ranking and are discarded.

Docking methods vary in the way sampling and scoring are implemented, and also in the representation of the molecules in the calculations. An important choice to be made is whether the proteins are kept rigid or whether flexibility is needed. Flexibility can be introduced in various ways, e.g. by using an ensemble of rigid structures (experimental or generated for example by molecular dynamics methods) corresponding to static snapshots of possible conformational changes, by allowing some interpenetration of the docked molecules (sometimes called ‘soft’ rigid body docking, as opposed to ‘hard’ rigid body docking, where no overlap is allowed at all), or by allowing explicit side-chain and/or backbone flexibility during the docking. The type of sampling depends on the way in which the molecules are represented. When a grid representation of the molecules is used, rigid body docking can be done by calculating correlations (e.g. surface complementarity) using fast Fourier transform methods [28–33]. When the protein is explicitly represented using an atomic model, one can use various sampling methods such as Monte Carlo [34–36] and molecular dynamics methods [7] or genetic algorithms [36] in combination with simulated annealing schemes. The scoring is typically based on some kind of force field [37], which assigns an energy to atom–atom (or

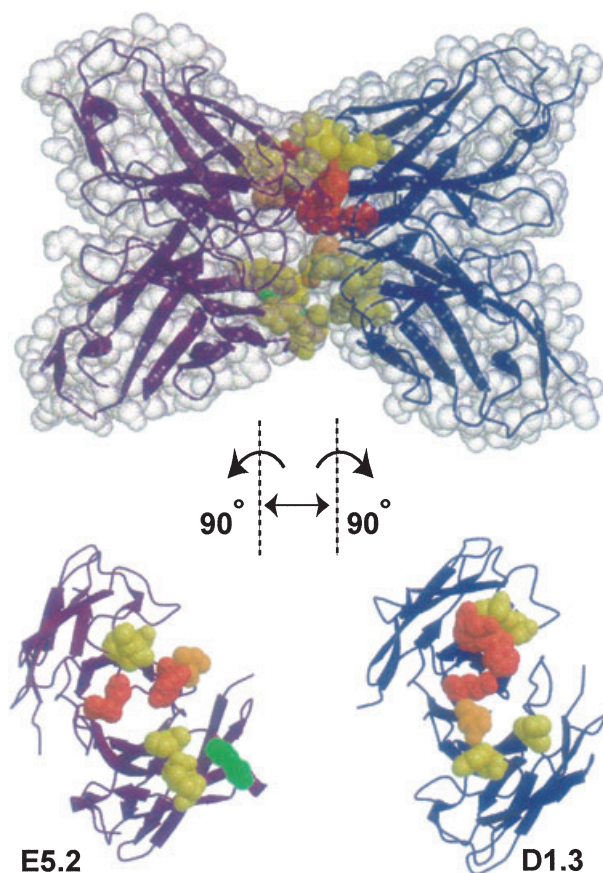


Fig. 2. Mapping of the mutagenesis data [177] on to the structure of the antibody D1.3–antibody E5.2 complex [187] (pdb entry 1dvf). Top: structure of the complex; bottom: interaction surface of E5.2 (left) and D1.3 (right) color coded according to the measured $\Delta\Delta G$ value [177] in mutagenesis experiments. Red: $\Delta\Delta G > 4.0$ kcal·mol⁻¹; orange: $\Delta\Delta G$ 2.1–4.0 kcal·mol⁻¹; yellow: $\Delta\Delta G$ 1.1–2.0 kcal·mol⁻¹; green: $\Delta\Delta G < 1.0$ kcal·mol⁻¹. Figures are prepared using MOLSCRIPT [188] and RASTER3D [189].

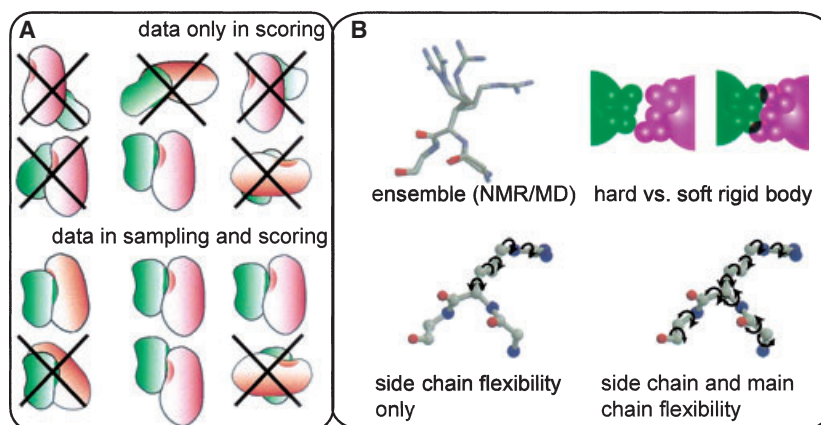
residue–residue) pairs, and subsequently adds all these together to get the energy of a given configuration. Often, terms such as buried surface area and

desolvation energy are added. Force fields can have a physical basis or can be knowledge based (derived by counting how often a given pair occurs in a database of experimental structures). Using biochemical and/or biophysical data in docking approaches has advantages for both the sampling and scoring stages. During the sampling, more ‘relevant’ configurations are produced, whereas in the scoring, the ranking of true positives (i.e. correct solutions) can be improved compared with *ab initio* docking, where typically tens to hundreds of false positives are scored at the top. An important difference between various methods is whether the experimental data are only introduced in the scoring (i.e. to filter the solutions that have been generated) or whether they are also used during sampling. In the following we will discuss a number of methods that have been proposed, first the procedures that only use experimental data for scoring, and next those that incorporate experimental data into the sampling itself. In Fig. 3 a graphical representation is given of the choices to make in the various docking approaches with respect to the incorporation of experimental data and the treatment of flexibility.

Although computer-based approaches should be preferred in terms of reproducibility, it is also possible to ‘manually’ build models of complexes based on experimental information. In fact there are quite a few examples where this has been done [38–42], some of which have been compared with pure *ab initio* docking results [43].

We should point out here that each docking approach has its own advantages and disadvantages, and the ‘docking problem’ is still unsolved: no single docking method will always give the right answer. The docking field is still in active development, and various approaches to the problem are being pursued, as will be discussed below.

Fig. 3. Some choices to be made in docking. (A) When to introduce the data? Here the complex structures resulting from a hypothetical docking method are shown, and the scoring is represented in a simplified way, discarding the complexes that do not satisfy the experimental restraints (indicated by the black crosses); (B) How to deal with flexibility: using an ensemble of starting structures; by soft rigid body docking; and explicitly during the docking by allowing side chain and/or main chain flexibility.



Docking methods using experimental data only in the scoring stage

A large variety of docking methods exist and have been used before applying a filter based on experimental data. One approach consists of a systematic grid search for all possible orientations (three translations and six rotations). This is only feasible for small systems and simplified models, as otherwise scoring all possible configurations becomes intractable. Such a method has been used for probing transmembrane helix multimers, e.g. the dimeric transmembrane region of glycoporphin A and the phospholamban pentamer. The low-energy structures resulting from the grid search were filtered using mutagenesis data [44–47].

When studying larger systems, and especially if one wants to introduce sophisticated amounts of flexibility in the docking, exhaustive grid searches become unrealistic. A fast method to perform grid calculations based on spherical Fourier correlations is implemented in the program Hex [48]. It has been combined with mutagenesis data [49]. Fast Fourier transform methods have often been used in docking. For example, the docking program DOT [29] has been used in combination with MS H/D data to filter solutions [50]. Other examples of fast Fourier transform based methods are the soft docking program GRAMM [30], which has been used in combination with mutagenesis data [51] and FTDOCK [28], which was originally tested on several complexes using experimental data (e.g. active-site information in the case of enzyme–inhibitor complexes) and was recently combined with NMR data (CSP and RDCs) to filter solutions [52]. Another grid approach, which uses Boolean-type operations and was optimized heuristically for speed, is the docking program BIGGER [53]. This program allows soft rigid body docking (hard and soft docking are compared in [54]). BIGGER is often used in combination with NMR CSP data [55–59].

There are several docking approaches that do not use a grid but rather an explicit search in the configurational space, e.g. DOCK [60,61], AUTODOCK [36], which was used in combination with CSP data [62], and other methods based on Brownian Dynamics simulations followed by Molecular Dynamic refinement of the initial models [63]. NMR CSP data have also been used in a more quantitative way for filtering docking solutions, by back-calculating chemical shift changes from the models with programs such as SHIFTS [64] or SHIFTX [65] and comparing them with the experimental values [66]. This approach has also been combined with RDCs [67]. The above methods have been successfully applied to model various biomolecular complexes (Tables 1–3).

Docking methods using experimental data to drive the docking

The advantage of using the data in the sampling stage of docking is that ‘correct’ or ‘near-correct’ configurations should be enriched, compared with approaches in which the data are only used in the scoring stage, provided of course that the experimental information is correct. This becomes especially important when the number of configurations is too large to be adequately sampled, as is often the case when flexibility is introduced.

As will be clear from the following discussion, there are different ways to incorporate the experimental data during the sampling stage. This partly depends on the kind of data used (e.g. the level of detail and the amount of inherent ambiguity) and the sampling method. ‘Geometric’ methods might limit the number of orientations selected for docking rather than adding experimental terms to an energy function. The search space is thus reduced on the basis of the available experimental data. The subsequent docking and scoring stages then proceed as in *ab initio* docking [68]. Other approaches use anchor points based on experimental data, e.g. TREEDOCK [69], or incorporate the experimental data by up weighting given residues in fast Fourier transform-based rigid body docking approaches (‘weighted geometric docking’) [32,70,71]. Another popular possibility is to use some kind of distance restraints. This means that an additional energy term is created, which is high if residues which, according to the data, should be at the interface, i.e. close to each other, are far away in the proposed complex, and, contrarily, low if they are near.

Ethylation interference and mutagenesis data have been used as experimental input for protein–DNA docking in the early data-driven Monte-Carlo docking program MONTY [34,72,73], which allows side-chain flexibility and DNA deformations. Double mutant cycle data, giving information about residue–residue contacts, have been incorporated as distance restraints in various applications [74–77]. A comparable approach was used to incorporate cross-linking data for a dimer of a four-transmembrane helix protein [78]: here a total of 10 distance restraints could be defined with quite small error bounds because of the rigid nature of the linker. There are several examples of the combination of NMR information with rigid body docking. Rigid body docking in X-PLOR [79] has been used to model the dynamic complex between plastocyanin and cytochrome *f* based on upper bound distance restraints derived from pseudo-contact shifts and CSP data, and lower bound distance restraints for residues assumed not to be in the interface [80,81]. Saturation transfer

and RDC restraints have been combined with energy minimization to model the CAD–ICAD complex (complex between the CAD domain of caspase activated deoxyribonuclease and the CAD domain of its inhibitor) [82]. The nucleoprotein superhelix–DNA complex was modeled using CSP restraints in a grid search [83].

Some experimental data are highly ambiguous and only provide information about interface residues, but not about the specific contacts they make. Docking approaches should thus be capable of incorporating such ambiguity. Typical examples here would be the CSP data obtained from NMR titration experiments or mutagenesis data. With this in mind, we developed an information-driven semiflexible docking approach called HADDOCK [7] in which any kind of information about interface residues can be incorporated as a highly ambiguous interaction restraint (AIR) (see below). Related approaches have been described in [84] where NMR CSP data and RDCs were used, and in [85] for cross-linking information detected by MS.

HADDOCK

The method

As is clear from the discussion above, there is a wealth of experimental sources that can provide information about interfaces of biomolecular complexes. These data are generally not used, however. Our docking approach HADDOCK, an acronym for high ambiguity driven docking [7], makes use of such information to drive the docking while allowing various degrees of flexibility. The information is encoded in AIRs similar to the ambiguous restraints commonly used in NMR structure determination [86]. The ambiguity here refers to the way in which the restraints are defined: between any residue which, based on experimental data, is believed to be an interface residue (called active residue), and all such residues (plus surface neighbors, called passive residues) on the partner molecule. An AIR is defined as an ambiguous intermolecular distance (d_{iAB}) 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_{res} in total) of protein B (n_{kB}) (and inversely for protein B). The effective distance d_{iAB}^{eff} for each restraint is calculated using the equation:

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

where N_{atoms} indicates all atoms of a given residue and N_{res} the sum of active and passive residues for a given

molecule. The definition of passive residues ensures that residues that are at the interface but are not detected (e.g. no CSP when using NMR, or no change in binding on mutation) are still able to satisfy the AIR restraints, i.e. contact active residues of the partner molecule. The $1/r^6$ summation [87] is used to mimic the attractive part of a Lennard-Jones potential and ensures that the AIRs are satisfied as soon as any two atoms of the two proteins are in contact. The AIRs are incorporated as an additional energy term to the energy function that one tries to minimize during the sampling. The docking proceeds in three stages during which increasing amounts of flexibility are introduced. In the first stage, the molecules are considered as rigid bodies, and a large number of solutions are generated. In the second stage, a limited amount of flexibility is introduced first into the side chains and subsequently into both side chains and backbone of predefined flexible segments encompassing the active and passive residues. Finally, the solutions are refined in explicit solvent. The final structures are clustered and scored using a combination of energy terms (mainly intermolecular van der Waals and electrostatic energies and restraint energies); for details see [7,88]. Note that fully flexible models can also be defined, for example for the docking of an unstructured peptide on to a protein.

Applications

Several groups have used HADDOCK to generate models of biomolecular complexes in combination with different sources of information such as mutagenesis [89–91] or NMR CSP data [88,89,91–96]. A common problem resulting from the highly ambiguous nature of the interaction restraints is that symmetrical solutions are often obtained corresponding, for example, to a 180° rotation of one molecule with respect to the other. In cases where energy considerations cannot distinguish between the symmetrical solutions, additional information should ideally be supplemented. This was the case for the UbcH5–Not4 complex [88] (Fig. 4A). To solve the symmetry problem, the HADDOCK models were used for structure-directed mutagenesis. Reverse mutants could be produced in which two residues of opposite charges across the interface were swapped, restoring thereby the binding. This provided unique, unambiguous information to select the correct solution.

In the case of the transient complex between the yeast copper chaperone Atx1 and the first soluble domain of the copper-transporting ATPase Cccp2, a copper ion was explicitly introduced into the docking calculations based on NMR CSP data and found to

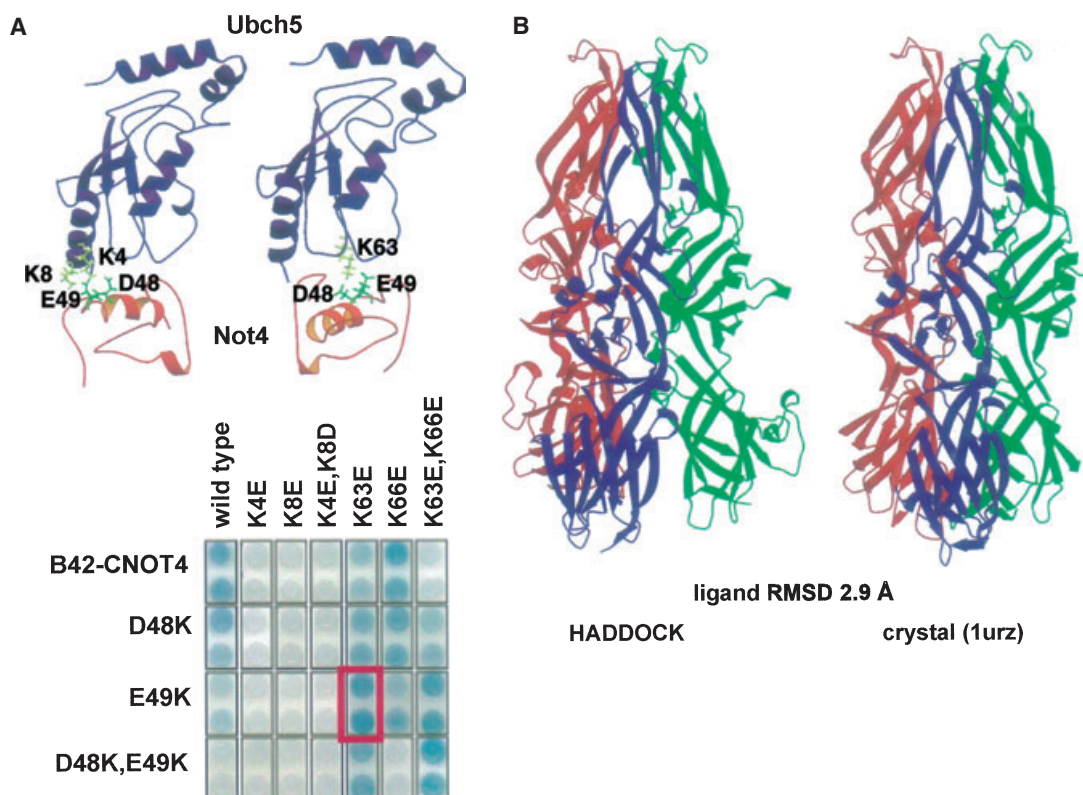


Fig. 4. Two examples of structures calculated using HADDOCK. (A) The Ubch5–Not4 complex (pdb entry 1ur6) [88]. In a first docking run using only NMR CSP data, two models were obtained (top left and top right). Based on these, mutagenesis experiments were performed to discriminate between the two models: the charge-reversing double mutant E49K,K63E did restore the complex (red box), whereas the double mutants including K4E or K8E did not restore complex formation. Only the left solution is consistent with this information. (B) TBE virus envelope glycoprotein E trimer (CAPRI target 10), for which epitope, conservation and protection from enzymatic digestion data were introduced in HADDOCK, resulting in a docking model (left) within 2.9 Å ligand–RMSD from the crystal structure [190] (pdb entry 1urz, right). The three subunits are color-coded; note that two segments (residue 148–159 and 204–209) are missing from the crystal structure.

move from Atx1 to Cccp2, consistent with the physiological direction of transfer [92]. The copper-transfer intermediate was a result of the flexible docking protocol, as no restraints were introduced to force the copper ion to move. This example indicates that flexible data-driven docking can be used to investigate not only ‘static’ structures but also more ‘dynamic’ aspects of biomolecular complexes. When available, classical NMR data such as NOEs can also be incorporated into HADDOCK, as was the case for generating the solution structure of a nonspecific protein–DNA complex [97].

Recently, we participated in the fourth and fifth round of the ‘blind docking competition’ CAPRI. As CAPRI is not especially meant for data-supported docking, we had to search literature and databases and use sequence conservation criteria (predicted via a neural network [98]) to define AIRs. Using HADDOCK, we were able to generate structures that are

close to the experimentally defined structures even with low-resolution, ‘fuzzy’ data such as epitope mapping and protection from enzymatic digestion. As an example, we successfully predicted the trimeric form of the TBE virus envelope glycoprotein E within 2.9 Å ligand–RMSD (Fig. 4B) (the ligand–RMSD is defined as the RMSD calculated on one component after superposition of the other components). Our participation in the CAPRI experiment has, however, taught us that in some cases our docking methods, as well as others, can fail.

Conclusions and perspectives

The combination of biochemical and biophysical data with docking has many different applications. Docking models can obviously be used to select residues to be targeted for mutagenesis, for example. One interesting point is that it becomes possible, when flexibility is

explicitly introduced, to investigate structural changes at the interface on complex formation, or even dynamic events as shown above for the copper-transfer complex. Here we discuss what the future of this kind of approach might be.

Perspectives on data used in docking

One interesting development is the use of conservation data to define interface residues (reviewed in [99]). Several methods have been developed for this purpose; examples are the use of a neural network [98,100], the determination of invariant polar residues [101], 3D cluster analysis [102], the use of phylogenetic trees, [103] the Evolutionary Trace method [104,105] and the Promate approach where conservation is combined with general interface characteristics [106]. Information from predicted interfaces has been used to model several complexes, for example, the Hsp90-p23 [107] and $G\alpha\beta\gamma$ trimer–receptor complexes [42] based on predictions obtained with the Evolutionary Trace method, and the complex between the $\alpha 1$ and $\beta 2$ subunits of hemoglobin and the FtsA homodimer [43] based on conservation data and correlated mutations [46]. With the increasing amount of genomic data available, this kind of analysis can be expected to become more and more important. In addition, protein interaction networks can be compared using PATHBLAST [108]; homologies based on this may provide additional information. Similarly, homology modeling, which has been improving over the years [109], in addition to being used to generate starting structures, could be combined with docking approaches, as illustrated with mutagenesis and neutron-scattering data [110] and MS data [111,112]. An interesting example of the combination of homology modeling and docking is the Multiprospector multimeric threading approach [113], which has been applied to the *Saccharomyces cerevisiae* proteome [114]: Multiprospector threads the sequences of the single chains of a target complex; if a template is found that is part of a complex, both chains of the target are rethreaded, now also incorporating an interfacial energy term.

Two experimental techniques which are very promising in combination with docking are cryo-electron microscopy or tomography and SAXS. Both techniques provide ‘shape’ information into which the structures of known constituents of a complex can be fitted. Cryo-electron microscopy has been used for a large number of yeast complexes [115] and for the 80S ribosome from *S. cerevisiae* [116]. For further discussion see reference [8]. SAXS data have been applied in docking to a variety of systems [117–124]. Specific

examples are the twinfilin-capping protein complex [125] for which models of the single components were fitted to the SAXS data and compared with mutagenesis data, and the FixJ response regulator where the rotation angle between the two domains was probed [126].

Another technique that can potentially be used is fluorescence. Interface information could be obtained for example for the complex of HscA with IscU LPPVK motif-containing peptides [127]: the ability of Trp residues at the N-terminus or C-terminus of the peptides to quench the fluorescence of labeled HscA was measured, and this allowed us to define the substrate-binding orientation. In another example, docking simulations of HLA-1 dimers and complexes of those with CD8 and TCR were compared with fluorescence resonance energy transfer data [128]. The use of fluorescence resonance energy transfer to study protein–DNA interactions has been reviewed [129]. Infrared spectroscopy might also become useful. For example, it was possible to define the tilt and relative orientation of transmembrane helices in the pentameric phospholamban [130] and the tetrameric M2 protein complex [131] based on infrared data.

With respect to the techniques discussed above, at least for MS and NMR, improvements can be expected. An example of a new MS approach for mapping interfaces is the modification of solvent-accessible side chains by hydroxyl radicals from millisecond exposure of aqueous solutions to X-rays; the modification sites can be identified by MS and differences between complexed and uncomplexed forms indicate the location of the binding interface [132,133]. In NMR, new approaches are emerging that might overcome the assignment problem. Comparison of experimental and back-calculated unassigned 1D ^1H spectra of a complex has been proposed as a means of filtering docking solutions; the feasibility of this approach has been demonstrated for four complexes [134]. Other methods that do not require chemical shift assignments but rely on the combination of amino acid-specific labeling with saturation transfer or titration experiments have been reported as well [135,136]. Provided that selective labeling can be efficiently performed, such methods should clearly speed up interface mapping by NMR.

Considering that information-driven docking will be much faster than conventional structural methods, it makes sense to invest some time and effort in making sure that the experimental data are reliable and really reveal interface residues. Therefore, whatever experimental technique is preferred, it is worth combining information from various sources.

Perspectives on docking methods

Not only from the data side, but also from the methodological point of view, improvements are needed and can be expected. It will be possible one day to perform reliable *ab initio* docking, in which case no data will be needed at all, but this is probably not within our reach for the coming years. Still, active developments in the *ab initio* docking field will definitely benefit data-driven docking approaches. Next to the need for proper scoring schemes, another important aspect is the handling of flexibility during docking. Although several methods exist that perform reasonably well in this respect, many still only use rigid body (soft) docking. Potential improvements might include a more widespread use of energy-driven sampling methods, such as molecular dynamics, before docking to generate ensembles of starting structures, during docking to allow induced conformational changes, and/or after docking to refine the (rigid body) solutions. Other advanced computational methods are emerging aiming at identifying parts of a molecule that are likely to be flexible and undergo conformational changes on complex formation [137,138]. Another kind of flexibility which, in our opinion without a good reason, has not had much attention is that complexes themselves might be dynamic. As the forces that hold together the non-covalently linked complexes are, in most cases, weaker than those that are involved in covalent interactions, one would expect mobility to play a bigger role here. This will be particularly true in the case of weak and transient complexes. Methods should be developed that take this into account.

Perspectives on experimental systems amenable to data-driven docking

Finally, the range of systems studied with docking approaches can also be extended. Although it might not strictly speaking be docking, it is interesting to note that the kind of methods that we have discussed here in the context of biomolecular complexes can also be applied to generate structures of single proteins by docking structural elements. This was done using cross-linking data to refine a homology model of FGF-2 [139] and with distance restraints for the lactose permease which consists of 12 transmembrane helices [140]. In another example, dipolar EPR distances, disulfide mapping distances and electron cryo-microscopy data were used in a special kind of exhaustive search using a graph-theory algorithm to generate models of rhodopsin [141]. Docking-like

approaches are particularly interesting for modeling transmembrane helical proteins, as these typically contain considerable helical content already in their unfolded state; this means that docking approaches can be applied using helical segments as structural entities, as described for example in reference [142]. A general review about helix-helix interactions in the folding of membrane protein can be found in reference [143].

At the other extreme, data have become available for many giant multisubunit complexes such as the ribosome [144] or the regulatory complex of the *Drosophila* 26S proteasome [145], but docking approaches have not often been used for them. A combinatorial approach such as CombDock [146] may be useful here, but HADDOCK or other docking methods can also easily be extended to deal with multiple subunits (as shown for the trimer example above), although, for large assemblies, computational requirements might become a limiting factor. Another kind of biological system for which data are becoming available now are protein-lipid assemblies. Using EPR, the orientation of phospholipase A₂ [147,148] with respect to the surface of phospholipid vesicles was studied. For the C2 domain of protein kinase A, fluorescence and EPR data were used to elucidate the surface of the protein that contacts the membrane and to generate a model for the protein attached to a membrane [149]. NMR spin label data have also been used to provide the depth and angle of micelle insertion of the FYVE domain of early endosome antigen I [150]. Finally, one interesting type of system to which increasing attention is given consists of proteins that, in their monomeric form, are unstructured and only fold during complex formation. A docking approach was used to study the complex of the (prefolded) actin with the (only folding upon binding) thymosin β 4, using a combination of NMR data, mutation data and cross-linking data as restraints in the docking [151].

In conclusion, we have shown that docking methods can provide valuable biological insight, when combined with a limited amount of experimental data. Such a combination will, without doubt, become more widely used in the near future.

Acknowledgements

Financial support from the Netherlands Organization for Scientific Research (N.W.O.) through a *Jonge Chemici* grant to A.M.J.J.B. (grant number 700.50.512) is acknowledged. We thank Cyril Dominguez and Sjoerd de Vries (Utrecht University) for helpful discussions.

References

- Halperin I, Ma BY, Wolfson H & Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* **47**, 409–443.
- Wodak SJ & Janin J (2003) Structural basis of macromolecular recognition. *Adv Protein Chem* **61**, 9–73.
- Brooijmans N & Kuntz ID (2003) Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **32**, 335–373.
- Vajda S & Camacho CJ (2004) Protein–protein docking: is the glass half-full or half-empty? *Trends Biotechnol* **22**, 110–116.
- Janin J, Henrick K, Moult J, Ten Eyck L, Sternberg MJE, Vajda S, Vasker I & Wodak SJ (2003) CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **52**, 2–9.
- Fradera X & Mestres J (2004) Guided docking approaches to structure-based design and screening. *Curr Top Med Chem* **4**, 687–700.
- Dominguez C, Boelens R & Bonvin AMJJ (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731–1737.
- Russell RB, Alber F, Aloy P, Davis FP, Korkin D, Pichaud M, Topf M & Sali A (2004) A structural perspective on protein–protein interactions. *Curr Opin Struct Biol* **14**, 313–324.
- McDonnell JM (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Curr Opin Chem Biol* **5**, 572–577.
- Vidal M, Brachmann RK, Fattaey A, Harlow E & Boeke JD (1996) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein–protein and DNA–protein interactions. *Proc Natl Acad Sci USA* **93**, 10315–10320.
- Sidhu SS, Fairbrother WJ & Deshayes K (2003) Exploring protein–protein interactions with phage display. *Chembiochem* **4**, 14–25.
- Clackson T & Wells JA (1995) A hot-spot of binding-energy in a hormone–receptor interface. *Science* **267**, 383–386.
- DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* **12**, 14–20.
- Thorn KS & Bogan AA (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **17**, 284–285.
- Carter PJ, Winter G, Wilkinson AJ & Fersht AR (1984) The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**, 835–840.
- Hanson CL & Robinson CV (2004) Protein–nucleic acid interactions and the expanding role of mass spectrometry. *J Biol Chem* **279**, 24907–24910.
- Hernandez H & Robinson CV (2001) Dynamic protein complexes: insights from mass spectrometry. *J Biol Chem* **276**, 46685–46688.
- Lanman J & Prevelige PE (2004) High-sensitivity mass spectrometry for imaging subunit interactions: hydrogen/deuterium exchange. *Curr Opin Struct Biol* **14**, 181–188.
- Garcia RA, Pantazatos D & Villarreal FJ (2004) Hydrogen/deuterium exchange mass spectrometry for investigating protein–ligand interactions. *Assay Drug Dev Technol* **2**, 81–91.
- Back JW, de Jong L, Muijsers AO & de Koster CG (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* **331**, 303–313.
- Zuiderweg ER (2002) Mapping protein–protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1–7.
- Takahashi H, Nakanishi T, Kami K, Arata Y & Shimada I (2000) A novel NMR method for determining the interfaces of large protein–protein complexes. *Nat Struct Biol* **7**, 220–223.
- Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci* **12**, 1–16.
- Fushman D, Varadan R, Assfalg M & Walker O (2004) Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. *Prog Nucl Magn Reson Spectrosc* **44**, 189–214.
- Gaponenko V, Altieri AS, Li J & Byrd RA (2002) Breaking symmetry in the structure determination of (large) symmetric protein dimers. *J Biomol NMR* **24**, 143–148.
- Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J & Byrd RA (2004) Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long-range restraints. *J Biomol NMR* **28**, 205–212.
- Arumugam S & Van Doren SR (2003) Global orientation of bound MMP-3 and N-TIMP-1 in solution via residual dipolar couplings. *Biochemistry* **42**, 7950–7958.
- Gabb HA, Jackson RM & Sternberg MJE (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* **272**, 106–120.
- Mandell JG, Roberts VA, Pique ME, Kotlovsky V, Mitchell JC, Nelson E, Tsigelny I & Ten Eyck LF (2001) Protein docking using continuum electrostatics and geometric fit. *Protein Eng* **14**, 105–113.
- Vakser IA (1995) Protein docking for low-resolution structures. *Protein Eng* **8**, 371–377.

- 31 Meyer M, Wilson P & Schomburg D (1996) Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. *J Mol Biol* **264**, 199–210.
- 32 Ben-Zeev E & Eisenstein M (2003) Weighted geometric docking: incorporating external information in the rotation-translation scan. *Proteins* **52**, 24–27.
- 33 Chen R, Li L & Weng ZP (2003) ZDOCK: An initial-stage protein-docking algorithm. *Proteins* **52**, 80–87.
- 34 Knegtel RMA, Boelens R & Kaptein R (1994) Monte Carlo docking of protein–DNA complexes: incorporation of DNA flexibility and experimental data. *Protein Eng* **7**, 761–767.
- 35 Abagyan R, Totrov M & Kuznetsov D (1994) Icm – a new method for protein modeling and design – applications to docking and structure prediction from the distorted native conformation. *J Comp Chem* **15**, 488–506.
- 36 Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK & Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comp Chem* **19**, 1639–1662.
- 37 Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. *J Comp Chem* **25**, 1584–1604.
- 38 Rajesh S, Sakamoto T, Iwamoto-Sugai M, Shibata T, Kohno T & Ito Y (1999) Ubiquitin binding interface mapping on yeast ubiquitin hydrolase by NMR. *Biochemistry* **38**, 9242–9253.
- 39 Singh S, Folkers GE, Bonvin AMJJ, Boelens R, Wechselberger R, Niztayev A & Kaptein R (2002) Solution structure and DNA-binding properties of the C-terminal domain of UvrC from *E. coli*. *EMBO J* **21**, 6257–6266.
- 40 Chien CY, Xu YJ, Xiao R, Aramini JM, Sahasrabudhe PV, Krug RM & Montelione GT (2004) Biophysical characterization of the complex between double-stranded RNA and the N-terminal domain of the NS1 protein from influenza A virus: evidence for a novel RNA-binding mode. *Biochemistry* **43**, 1950–1962.
- 41 Comolli LR, Pelton JG & Tinoco I (1998) Mapping of a protein–RNA kissing hairpin interface: Rom and Tar-Tar. *Nucleic Acids Res* **26**, 4688–4695.
- 42 Lichtarge O, Bourne HR & Cohen FE (1996) Evolutionarily conserved G (alpha beta gamma) binding surfaces support a model of the G protein–receptor complex. *Proc Natl Acad Sci USA* **93**, 7507–7511.
- 43 Carettoni D, Gomez-Puertas P, Yim L, Mingorance J, Massidda O, Vicente M, Valencia A, Domenici E & Anderluzzi D (2003) Phage-display and correlated mutations identify an essential region of subdomain 1C involved in homodimerization of *Escherichia coli* FtsA. *Proteins* **50**, 192–206.
- 44 Adams PD, Arkin IT, Engelman DM & Brunger AT (1995) Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nat Struct Biol* **2**, 154–162.
- 45 Adams PD, Engelman DM & Brunger AT (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* **26**, 257–261.
- 46 Pazos F, Helmer-Citterich M, Ausiello G & Valencia A (1997) Correlated mutations contain information about protein–protein interaction. *J Mol Biol* **271**, 511–523.
- 47 Li RH, Gorelik R, Nanda V, Law PB, Lear JD, DeG-rado WF & Bennett JS (2004) Dimerization of the transmembrane domain of integrin alpha (IIb) subunit in cell membranes. *J Biol Chem* **279**, 26666–26673.
- 48 Ritchie DW & Kemp GJL (2000) Protein docking using spherical polar Fourier correlations. *Proteins* **39**, 178–194.
- 49 Gaboriaud C, Juanhuix J, Gruez A, Lacroix M, Darnault C, Pignol D, Verger D, Fontecilla-Camps JC & Arlaud GJ (2003) The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J Biol Chem* **278**, 46974–46982.
- 50 Anand GS, Law D, Mandell JG, Snead AN, Tsigelny I, Taylor SS, Ten Eyck LF & Komives EA (2003) Identification of the protein kinase A regulatory R-I alpha–catalytic subunit interface by amide H/H-2 exchange and protein docking. *Proc Natl Acad Sci USA* **100**, 13264–13269.
- 51 Azuma Y, Renault L, Garcia-Ranea JA, Valencia A, Nishimoto T & Wittinghofer A (1999) Model of the Ran–RCC1 interaction using biochemical and docking experiments. *J Mol Biol* **289**, 1119–1130.
- 52 Dobrodumov A & Gronenborn AM (2003) Filtering and selection of structural models: combining docking and NMR. *Proteins* **53**, 18–32.
- 53 Palma PN, Krippahl L, Wampler JE & Moura JGG (2000) BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* **39**, 372–384.
- 54 Pettigrew GW, Pauleta SR, Goodhew CF, Cooper A, Nutley M, Jumel K, Harding SE, Costa C, Krippahl L, Moura I & Moura J (2003) Electron transfer complexes of cytochrome c peroxidase from *Paracoccus denitrificans* containing more than one cytochrome. *Biochemistry* **42**, 11968–11981.
- 55 Morelli XJ, Palma PN, Guerlesquin F & Rigby AC (2001) A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data. *Protein Sci* **10**, 2131–2137.
- 56 Crowley PB, Rabe KS, Worrall JAR, Canters GW & Ubbink M (2002) The ternary complex of cytochrome f and cytochrome c: identification of a second binding site and competition for plastocyanin binding. *Chem-biochem* **3**, 526–533.
- 57 Worrall JAR, Liu YJ, Crowley PB, Nocek JM, Hoffman BM & Ubbink M (2002) Myoglobin and

- cytochrome b (5): a nuclear magnetic resonance study of a highly dynamic protein complex. *Biochemistry* **41**, 11721–11730.
- 58 Morelli X, Dolla A, Czjzek M, Palma PN, Blasco F, Krippahl L, Moura JIG & Guerlesquin F (2000) Heteronuclear NMR and soft docking: an experimental approach for a structural model of the cytochrome c (553)–ferredoxin complex. *Biochemistry* **39**, 2530–2537.
- 59 McKenna S, Moraes T, Pastushok L, Ptak C, Xiao W, Spyropoulos L & Ellison MJ (2003) An NMR-based model of the ubiquitin-bound human ubiquitin conjugation complex Mms2-Ubc13: the structural basis for lysine 63 chain catalysis. *J Biol Chem* **278**, 13151–13158.
- 60 Meng EC, Gschwend DA, Blaney JM & Kuntz ID (1993) Orientational sampling and rigid-body minimization in molecular docking. *Proteins* **17**, 266–278.
- 61 Cuff L, Ulrich RG & Olson MA (2003) Prediction of the multimeric assembly of staphylococcal enterotoxin A with cell-surface protein receptors. *J Mol Graph Model* **21**, 473–486.
- 62 Sachchidanand Lequin O, Staunton D, Mulloy B, Forster MJ, Yoshida K & Campbell ID (2002) Mapping the heparin-binding site on the (13–14), F3 fragment of fibronectin. *J Biol Chem* **277**, 50629–50635.
- 63 Yu, K, Fu W, Liu H, Luo X, Chen KX, Ding J, Shen J & Jiang H (2004) Computational simulations of interactions of scorpion toxins with the voltage-gated potassium ion channel. *Biophys J* **86**, 3542–3555.
- 64 Xu XP & Case DA (2001) Automated prediction of N-15, C-13 (alpha), C-13 (beta) and C-13 'chemical shifts in proteins using a density functional database. *J Biomol NMR* **21**, 321–333.
- 65 Neal S, Nip AM, Zhang HY & Wishart DS (2003) Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. *J Biomol NMR* **26**, 215–240.
- 66 Stamos J, Eigenbrot C, Nakamura GR, Reynolds ME, Yin J, Lowman HB, Fairbrother WJ & Starovasnik MA (2004) Convergent recognition of the IgE binding site on the high-affinity IgE receptor. *Structure* **12**, 1289–1301.
- 67 McCoy MA & Wyss DF (2002) Structures of protein–protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. *J Am Chem Soc* **124**, 2104–2105.
- 68 Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R & Wolfson HJ (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* **52**, 107–112.
- 69 Fahmy A & Wagner G (2002) TreeDock: a tool for protein docking based on minimizing van der Waals energies. *J Am Chem Soc* **124**, 1241–1250.
- 70 Ben-Zeev E, Zarivach R, Shoham M, Yonath A & Eisenstein M (2003) Prediction of the structure of the complex between the 30S ribosomal subunit and colicin E3 via weighted-geometric docking. *J Biomol Struct Dyn* **20**, 669–675.
- 71 Zarivach R, Ben-Zeev E, Wu N, Auerbach T, Bashan A, Jakes K, Dickman K, Kosmidis A, Schlutzen F, Yonath A, Eisenstein M & Shoham M (2002) On the interaction of colicin E3 with the ribosome. *Biochimie* **84**, 447–454.
- 72 Knegtel RMA, Fogh RH, Otteleben G, Ruterjans H, Dumoulin P, Schnarr M, Boelens R & Kaptein R (1995) A model for the Lexa repressor DNA complex. *Proteins* **21**, 226–236.
- 73 Rumpel S, Razeto A, Pillar CM, Vijayan V, Taylor A, Giller K, Gilmore MS, Becker S & Zweckstetter M (2004) Structure and DNA-binding properties of the cytolysin regulator CylR2 from *Enterococcus faecalis*. *EMBO J* **23**, 3632–3642.
- 74 Gilquin B, Racape J, Wrisch A, Visan V, Lecoq A, Grissmer S, Menez A & Gasparini S (2002) Structure of the BgK-Kv1.1 complex based on distance restraints identified by double mutant cycles: molecular basis for convergent evolution of Kv1 channel blockers. *J Biol Chem* **277**, 37406–37413.
- 75 Eriksson MAL & Roux B (2002) Modeling the structure of Agitoxin in complex with the Shaker K⁺ channel: a computational approach based on experimental distance restraints extracted from thermodynamic mutant cycles. *Biophys J* **83**, 2595–2609.
- 76 Fruchart-Gaillard C, Gilquin B, Antil-Delbeke S, Le Novère N, Tamiya T, Corringer PJ, Changeux JP, Menez A & Servent D (2002) Experimentally based model of a complex between a snake toxin and the alpha 7 nicotinic receptor. *Proc Natl Acad Sci USA* **99**, 3216–3221.
- 77 Roisman LC, Piehler J, Trosset JY, Scheraga HA & Schreiber G (2001) Structure of the interferon–receptor complex determined by distance constraints from double-mutant cycles and flexible docking. *Proc Natl Acad Sci USA* **98**, 13231–13236.
- 78 Gottschalk K-E, Soskine M, Schuldiner S & Kessler H (2004) A structural model of EmrE, a multi-drug transporter from *Escherichia coli*. *Biophys J* **86**, 3335–3348.
- 79 Brunger AT (1992) X-PLOR 3.1 Manual. Yale University Press, New Haven, CT, USA.
- 80 Ubbink M, Ejdeback M, Karlsson BG & Bendall DS (1998) The structure of the complex of plastocyanin and cytochrome f, determined by paramagnetic NMR and restrained rigid-body molecular dynamics. *Structure* **6**, 323–335.
- 81 Crowley PB, Otting G, Schlarb-Ridley BG, Canters GW & Ubbink M (2001) Hydrophobic interactions in a cyanobacterial plastocyanin–cytochrome f complex. *J Am Chem Soc* **123**, 10444–10453.
- 82 Matsuda T, Ikegami T, Nakajima N, Yamazaki T & Nakamura H (2004) Model building of a protein–

- protein complexed structure using saturation transfer and residual dipolar coupling without paired intermolecular NOE. *J Biomol NMR* **29**, 325–338.
- 83 Folmer RHA, Nilges M, Papavoine CHM, Harmsen BJM, Konings RNH & Hilbers CW (1997) Refined structure, DNA binding studies, and dynamics of the bacteriophage Pf3 encoded single-stranded DNA binding protein. *Biochemistry* **36**, 9120–9135.
 - 84 Clore GM & Schwieters CD (2003) Docking of protein–protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from H-1 (N) /N-15 chemical shift mapping and backbone N-15-H-1 residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *J Am Chem Soc* **125**, 2902–2912.
 - 85 Schulz DM, Ihling C, Clore GM & Sinz A (2004) Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin–melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes. *Biochemistry* **43**, 4703–4715.
 - 86 Nilges M & O'Donoghue SI (1998) Ambiguous NOEs and automated NOE assignment. *Prog Nucl Magn Reson Spectrosc* **32**, 107–139.
 - 87 Nilges M (1993) A calculation strategy for the structure determination of symmetrical dimers by H-1-NMR. *Proteins* **17**, 297–309.
 - 88 Dominguez C, Bonvin AMJJ, Winkler GS, van Schaik FMA, Timmers HTM & Boelens R (2004) Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. *Structure* **12**, 633–644.
 - 89 Gao G, Prutzman KC, King ML, Scheswohl DM, DeRose EF, London RE, Schaller MD & Campbell SL (2004) NMR Solution structure of the focal adhesion targeting domain of focal adhesion kinase in complex with a paxillin LD peptide: evidence for a two-site binding model. *J Biol Chem* **279**, 8441–8451.
 - 90 Simpson RJY, Lee SHY, Bartle N, Sum EY, Visvader JE, Matthews JM, Mackay JP & Crossley M (2004) A classic zinc finger from friend of GATA mediates an interaction with the coiled-coil of transforming acidic coiled-coil 3. *J Biol Chem* **279**, 39789–39797.
 - 91 Jain NU, Wyckoff TJO, Raetz CRH & Prestegard JH (2004) Rapid analysis of large protein–protein complexes using NMR-derived orientational constraints: the 95 kDa complex of LpxA with Acyl carrier protein. *J Mol Biol* **343**, 1379–1389.
 - 92 Arnesano F, Banci L, Bertini I & Bonvin AMJJ (2004) A docking approach to the study of copper trafficking proteins: interaction between metallochaperones and soluble domains of copper ATPases. *Structure* **12**, 669–676.
 - 93 Mueller TD, Kamionka M & Feigon J (2004) Specificity of the interaction between ubiquitin-associated domains and ubiquitin. *J Biol Chem* **279**, 11926–11936.
 - 94 Stauffer ME & Chazin WJ (2004) Physical interaction between replication protein A and Rad51 promotes exchange on single-stranded DNA. *J Biol Chem* **279**, 25638–25645.
 - 95 van Drogen-Petit A, Zwahlen C, Peter M & Bonvin AM (2004) Insight into molecular interactions between two PB1 domains. *J Mol Biol* **336**, 1195–1210.
 - 96 Yuan XM, Simpson P, Mckeown C, Kondo H, Uchiyama K, Wallis R, Dreveny I, Keetch C, Zhang XD, Robinson C, Freemont P & Matthews S (2004) Structure, dynamics and interactions of p47, a major adaptor of the AAA ATPase, p97. *EMBO J* **23**, 1463–1473.
 - 97 Kalodimos CG, Biris N, Bonvin AMJJ, Levandoski MM, Guennegues M, Boelens R & Kaptein R (2004) Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science* **305**, 386–389.
 - 98 Zhou HX & Shan YB (2001) Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* **44**, 336–343.
 - 99 Lichtarge O & Sowa ME (2002) Evolutionary predictions of binding surfaces and interactions. *Curr Opin Struct Biol* **12**, 21–27.
 - 100 Fariselli P, Pazos F, Valencia A & Casadio R (2002) Prediction of protein–protein interaction sites in heterocomplexes with neural networks. *Eur J Biochem* **269**, 1356–1361.
 - 101 Aloy P, Querol E, Aviles FX & Sternberg MJE (2001) Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. *J Mol Biol* **311**, 395–408.
 - 102 Landgraf R, Xenarios I & Eisenberg D (2001) Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. *J Mol Biol* **307**, 1487–1502.
 - 103 Armon A, Graur D & Ben-Tal N (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J Mol Biol* **307**, 447–463.
 - 104 Lichtarge O, Bourne HR & Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* **257**, 342–358.
 - 105 Madabushi S, Yao H, Marsh M, Kristensen DM, Philippi A, Sowa ME & Lichtarge O (2002) Structural clusters of evolutionary trace residues are statistically significant and common in proteins. *J Mol Biol* **316**, 139–154.
 - 106 Neuvirth H, Raz R & Schreiber G (2004) ProMate: a structure based prediction program to identify the location of protein–protein binding sites. *J Mol Biol* **338**, 181–199.
 - 107 Zhu S & Tytgat J (2004) Evolutionary epitopes of Hsp90 and p23: implications for their interaction. *FASEB J* **18**, 940–947.

- 108 Kelley BP, Yuan BB, Lewitter F, Sharan R, Stockwell BR & Ideker T (2004) PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res* **32**, W83–W88.
- 109 Venclovas C, Zemla A, Fidelis K & Moulton J (2003) Assessment of progress over the CASP experiments. *Proteins* **53**, 585–595.
- 110 Tung CS, Walsh DA & Trewheella J (2002) A structural model of the catalytic subunit–regulatory subunit dimeric complex of the cAMP-dependent protein kinase. *J Biol Chem* **277**, 12423–12431.
- 111 D'Ambrosio C, Talamo F, Vitale RM, Amodeo P, Tell G, Ferrara L & Scaloni A (2003) Probing the dimeric structure of porcine aminoacylase 1 by mass spectrometric and modeling procedures. *Biochemistry* **42**, 4430–4443.
- 112 Taverner T, Hall NE, O'Hair RAJ & Simpson RJ (2002) Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J Biol Chem* **277**, 46487–46492.
- 113 Lu L, Lu H & Skolnick J (2002) Multiprospector: an algorithm for the prediction of protein–protein interactions by multimeric threading. *Proteins* **49**, 350–364.
- 114 Lu L, Arakaki AK, Lu H & Skolnick J (2003) Multimeric threading-based prediction of protein–protein interactions on a genomic scale: application to the *Saccharomyces cerevisiae* proteome. *Genome Res* **13**, 1146–1154.
- 115 Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin AC, Bork P, Superti-Furga G, Serrano L & Russell RB (2004) Structure-based assembly of protein complexes in yeast. *Science* **303**, 2026–2029.
- 116 Spahn CMT, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G & Frank J (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*: tRNA–ribosome and subunit–subunit interactions. *Cell* **107**, 373–386.
- 117 Dainese E, Svergun D, Beltramini M, Di Muro P & Salvato B (2000) Low-resolution structure of the proteolytic fragments of the *Rapana venosa* hemocyanin in solution. *Arch Biochem Biophys* **373**, 154–162.
- 118 de Azevedo WF, dos Santos GC, dos Santos DM, Olivieri JR, Canduri F, Silva RG, Basso LA, Renard G, da Fonseca IO, Mendes MA, Palma MS & Santos DS (2003) Docking and small angle X-ray scattering studies of purine nucleoside phosphorylase. *Biochem Biophys Res Commun* **309**, 923–928.
- 119 Grossmann JG, Sharff AJ, O'Hare P & Luisi B (2001) Molecular shapes of transcription factors TFIIB and VP16 in solution: implications for recognition. *Biochemistry* **40**, 6267–6274.
- 120 Svergun DI, Aldag I, Sieck T, Altendorf K, Koch MHJ, Kane DJ, Kozin MB & Gruber G (1998) A model of the quaternary structure of the *Escherichia coli* F-1 ATPase from X-ray solution scattering and evidence for structural changes in the delta subunit during ATP hydrolysis. *Biophys J* **75**, 2212–2219.
- 121 Callaghan AJ, Grossmann JG, Redko YU, Ilag LL, Moncrieffe MC, Symmons MF, Robinson CV, McDowell KJ & Luisi BF (2003) Quaternary structure and catalytic activity of the *Escherichia coli* ribonuclease E amino-terminal catalytic domain. *Biochemistry* **42**, 13848–13855.
- 122 Marquez JA, Smith CIE, Petoukhov MV, Lo Surdo P, Mattsson PT, Knekt M, Westlund A, Scheffzek K, Saraste M & Svergun DI (2003) Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering. *EMBO J* **22**, 4616–4624.
- 123 Auguin D, Barthe P, Royer C, Stern MH, Noguchi M, Arold ST & Roumestand C (2004) Structural basis for the co-activation of protein kinase B by T-cell leukemia-1 (TCL1) family proto-oncoproteins. *J Biol Chem* **279**, 35890–35902.
- 124 Sun Z, Reid KBM & Perkins SJ (2004) The dimeric and trimeric solution structures of the multidomain complement protein properdin by X-ray scattering, analytical ultracentrifugation and constrained modeling. *J Mol Biol* **343**, 1327–1343.
- 125 Falck S, Paavilainen VO, Wear MA, Grossmann JG, Cooper JA & Lappalainen P (2004) Biological role and structural mechanism of twinfilin-capping protein. *EMBO J* **23**, 3010–3019.
- 126 Birck C, Malfois M, Svergun D & Samama JP (2002) Insights into signal transduction revealed by the low resolution structure of the FixJ response regulator. *J Mol Biol* **321**, 447–457.
- 127 Tapley TL & Vickery LE (2004) Preferential substrate binding orientation by the molecular chaperone HscA. *J Biol Chem* **279**, 28435–28442.
- 128 Gaspar R, Bagossi P, Bene L, Matko J, Szollosi J, Tozser J, Fesus L, Waldmann TA & Damjanovich S (2001) Clustering of class IHLA oligomers with CD8 and TCR: Three-dimensional models based on fluorescence resonance energy transfer and crystallographic data. *J Immunol* **166**, 5078–5086.
- 129 Hillisch A, Lorenz M & Diekmann S (2001) Recent advances in FRET: distance determination in protein–DNA complexes. *Curr Opin Struct Biol* **11**, 201–207.
- 130 Torres J, Adams PD & Arkin IT (2000) Use of a new label C-13=O-18 in the determination of a structural model of phospholamban in a lipid bilayer. Spatial restraints resolve the ambiguity arising from interpretations of mutagenesis data. *J Mol Biol* **300**, 677–685.
- 131 Kukol A, Adams PD, Rice LM, Brunger AT & Arkin IT (1999) Experimentally based orientational refinement of membrane protein models: a structure for the Influenza A M2 H⁺ channel. *J Mol Biol* **286**, 951–962.
- 132 Guan JQ, Almo SC, Reisler E & Chance MR (2003) Structural reorganization of proteins revealed by

- radiolysis and mass spectrometry: G-actin solution structure is divalent cation dependent. *Biochemistry* **42**, 11992–12000.
- 133 Guan JQ, Almo SC & Chance MR (2004) Synchrotron radiolysis and mass spectrometry: a new approach to research on the actin cytoskeleton. *Acc Chem Res* **37**, 221–229.
 - 134 Kohlbacher O, Burchardt A, Moll A, Hildebrandt A, Bayer P & Lenhof HP (2001) Structure prediction of protein complexes by an NMR-based protein docking algorithm. *J Biomol NMR* **20**, 15–21.
 - 135 Hajduk PJ, Mack JC, Olejniczak ET, Park C, Dandliker PJ & Beutel BA (2004) SOS-NMR: a saturation transfer NMR-based method for determining the structures of protein-ligand complexes. *J Am Chem Soc* **126**, 2390–2398.
 - 136 Parker MJ, Aulton-Jones M, Hounslow AM & Craven CJ (2004) A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. *J Am Chem Soc* **126**, 5020–5021.
 - 137 Zacharias M (2004) Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: Binding of FK506 to FKBP. *Proteins: Structure Function Bioinformatics* **54**, 759–767.
 - 138 Kovacs JA, Chacon P & Abagyan R (2004) Predictions of protein flexibility: first-order measures. *Proteins: Structure Function Bioinformatics* **56**, 661–668.
 - 139 Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW & Dollinger G (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci USA* **97**, 5802–5806.
 - 140 Sorgen PL, Hu YL, Guan L, Kaback HR & Girvin ME (2002) An approach to membrane protein structure without crystals. *Proc Natl Acad Sci USA* **99**, 14037–14040.
 - 141 Faulon JL, Sale K & Young M (2003) Exploring the conformational space of membrane protein folds matching distance constraints. *Protein Sci* **12**, 1750–1761.
 - 142 Sale K, Faulon J-L, Gray GA, Schoeniger JS & Young MM (2004) Optimal bundling of transmembrane helices using sparse distance constraints. *Protein Sci* **13**, 2613–2627.
 - 143 DeGrado WF, Gratkowski H & Lear JD (2003) How do helix–helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci* **12**, 647–665.
 - 144 Whirl-Carrillo M, Gabashvili IS, Bada M, Banatao DR & Altman RB (2002) Mining biochemical information: lessons taught by the ribosome. *RNA* **8**, 279–289.
 - 145 Kurucz E, Ando I, Sumegi M, Holzl H, Kapelari B, Baumeister W & Udvardy A (2002) Assembly of the *Drosophila* 26S proteasome is accompanied by extensive subunit rearrangements. *Biochem J* **365**, 527–536.
 - 146 Inbar Y, Benyamini H, Nussinov R & Wolfson HJ (2003) Protein structure prediction via combinatorial assembly of sub-structural units. *Bioinformatics* **19**, 158i–168.
 - 147 Ball A, Nielsen R, Gelb MH & Robinson BH (1999) Interfacial membrane docking of cytosolic phospholipase A 2, C2 domain using electrostatic potential-modulated spin relaxation magnetic resonance. *Proc Natl Acad Sci USA* **96**, 6637–6642.
 - 148 Lin Y, Nielsen R, Murray D, Hubbell WL, Mailer C, Robinson BH & Gelb MH (1998) Docking phospholipase A (2) on membranes using electrostatic potential-modulated spin relaxation magnetic resonance. *Science* **279**, 1925–1929.
 - 149 Kohout SC, Corbalan-Garcia S, Gomez-Fernandez JC & Falke JJ (2003) C2 domain of protein kinase C alpha: elucidation of the membrane docking surface by site-directed fluorescence and spin labeling. *Biochemistry* **42**, 1254–1265.
 - 150 Kutateladze TG, Capelluto DGS, Ferguson CG, Cheever ML, Kutateladze AG, Prestwich GD & Overduin M (2004) Multivalent mechanism of membrane insertion by the FYVE domain. *J Biol Chem* **279**, 3050–3057.
 - 151 Domanski M, Hertzog M, Coutant J, Gutsche-Perelroizen I, Bontems F, Carlier MF, Guittet E & van Heijenoort C (2004) Coupling of folding and binding of thymosin beta 4 upon interaction with monomeric actin monitored by nuclear magnetic resonance. *J Biol Chem* **279**, 23637–23645.
 - 152 Norledge BV, Petrovan RJ, Ruf W & Olson AJ (2003) The tissue factor/factor VIIa/factor Xa complex: a model built by docking. *Proteins* **53**, 640–648.
 - 153 Sadir R, Baleux F, Grosdidier A, Imberty A & Lortat-Jacob H (2001) Characterization of the stromal cell-derived factor-1alpha-heparin. *J Biol Chem* **276**, 8288–8296.
 - 154 Karim CB, Stamm JD, Karim J, Jones LR & Thomas DD (1998) Cysteine reactivity and oligomeric structures of phospholamban and its mutants. *Biochemistry* **37**, 12074–12081.
 - 155 Herzyk P & Hubbard RE (1998) Using experimental information to produce a model of the transmembrane domain of the ion channel phospholamban. *Biophys J* **74**, 1203–1214.
 - 156 Jespers L, Lijnen HR, Vanwetswinkel S, Van Hoef B, Brepols K, Collen D & De Maeyer M (1999) Guiding a docking mode by phage display: selection of correlated mutations. *J Mol Biol* **290**, 471–479.
 - 157 Onrust R, Herzmark P, Chi P, Garcia PD, Lichtarge O, Kingsley C & Bourne HR (1997) Receptor and beta

- gamma binding sites in the alpha subunit of the retinal G protein transducin. *Science* **275**, 381–384.
- 158 Gruschus JM, Greene LE, Eisenberg E & Ferretti JA (2004) Experimentally biased model structure of the Hsc70/auxilin complex: substrate transfer and interdomain structural change. *Protein Sci* **13**, 2029–2044.
 - 159 Bracci L, Pini A, Bernini A, Lelli B, Ricci C, Scarselli M, Niccolai N & Neri P (2003) Biochemical filtering of a protein-protein docking simulation identifies the structure of a complex between a recombinant antibody fragment and alpha-bungarotoxin. *Biochem J* **371**, 423–427.
 - 160 Morillas M, Gomez-Puertas P, Rubi B, Clotet J, Arino J, Valencia A, Hegardt FG, Serra D & Asins G (2002) Structural model of a malonyl-CoA-binding site of carnitine octanoyltransferase and carnitine palmitoyltransferase I: mutational analysis of a malonyl-CoA affinity domain. *J Biol Chem* **277**, 11473–11480.
 - 161 Dumoulin P, Ebright RH, Knegtel R, Kaptein R, Granger-Schnarr M & Schnarr M (1996) Structure of the LexA repressor–DNA complex probed by affinity cleavage and affinity photo-cross-linking. *Biochemistry* **35**, 4279–4286.
 - 162 Aloy P, Moont G, Gabb HA, Querol E, Aviles FX & Sternberg MJE (1998) Modelling repressor proteins docking to DNA. *Proteins* **33**, 535–549.
 - 163 Tzou WS & Hwang MJ (1999) Modeling helix-turn-helix protein-induced DNA bending with knowledge-based distance restraints. *Biophys J* **77**, 1191–1205.
 - 164 Cai SJ, Khorchid A, Ikura M & Inouye M (2003) Probing catalytically essential domain orientation in histidine kinase EnvZ by targeted disulfide crosslinking. *J Mol Biol* **328**, 409–418.
 - 165 Dmitriev OY, Jones PC & Fillingame RH (1999) Structure of the subunit c oligomer in the F1F0 ATP synthase: model derived from solution structure of the monomer and cross-linking in the native enzyme. *Proc Natl Acad Sci USA* **96**, 7785–7790.
 - 166 You L, Gillilan R & Huffaker TC (2004) Model for the yeast cofactor A-beta-tubulin complex based on computational docking and mutagenesis. *J Mol Biol* **341**, 1343–1354.
 - 167 Lacroix M, Rossi V, Gaboriaud C, Chevallier S, Jaquinod M, Thielens NM, Gagnon J & Arlaud GJ (1997) Structure and assembly of the catalytic region of human complement protease C1r: a three-dimensional model based on chemical cross-linking and homology modeling. *Biochemistry* **36**, 6270–6282.
 - 168 Walters KJ, Lech PJ, Goh AM, Wang Q & Howley PM (2003) DNA-repair protein hHR23a alters its protein structure upon binding. *Proc Natl Acad Sci USA* **100**, 12694–12699.
 - 169 Varadan R, Walker O, Pickart C & Fushman D (2002) Structural properties of polyubiquitin chains in solution. *J Mol Biol* **324**, 637–647.
 - 170 Varadan R, Assfalg N, Haririnia A, Raasi S, Pickart C & Fushman D (2004) Solution conformation of Lys (63)-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *J Biol Chem* **279**, 7055–7063.
 - 171 Owen D, Lowe PN, Nietlispach D, Brosnan CE, Chirgadze DY, Parker PJ, Blundell TL & Mott HR (2003) Molecular dissection of the interaction between the small G proteins Rac1 and RhoA and protein kinase C-related kinase 1 (PRK1). *J Biol Chem* **278**, 50578–50587.
 - 172 McDonnell JM, Calvert R, Beavil RL, Beavil AJ, Henry AJ, Sutton BJ, Gould HJ & Cowburn D (2001) The structure of the IgE C epsilon 2 domain and its role in stabilizing the complex with its high-affinity receptor Fc epsilon RI alpha. *Nat Struct Biol* **8**, 437–441.
 - 173 Johnson MA & Pinto BM (2002) Saturation transfer difference 1D-TOCSY experiments to map the topography of oligosaccharides recognized by a monoclonal antibody directed against the cell-wall polysaccharide of Group A *Streptococcus*. *J Am Chem Soc* **124**, 15368–15374.
 - 174 Moller H, Serttas N, Paulsen H, Burchell JM, Taylor-Papadimitriou J & Meyer B (2002) NMR-based determination of the binding epitope and conformational analysis of MUC-1 glycopeptides and peptides bound to the breast cancer-selective monoclonal antibody SM3. *Eur J Biochem* **269**, 1444–1455.
 - 175 Buchko GW, Tung CS, McAteer K, Isern NG, Spicer LD & Kennedy MA (2001) DNA–XPA interactions: a P-31 NMR and molecular modeling study of dCCAATAACC association with the minimal DNA-binding domain (M98–F219) of the nucleotide excision repair protein XPA. *Nucleic Acids Res* **29**, 2635–2643.
 - 176 Schreiber G & Fersht AR (1995) Energetics of protein–protein interactions: analysis of the barnase–barstar interface by single mutations and double mutant cycles. *J Mol Biol* **248**, 478–486.
 - 177 Goldman ER, Dall, Acqua W, Braden BC & Mariuzza RA (1997) Analysis of binding interactions in an idiotope–antiidiotope protein–protein complex by double mutant cycles. *Biochemistry* **36**, 49–56.
 - 178 Pielak GJ & Wang X (2001) Interactions between yeast iso-1-cytochrome c and its peroxidase. *Biochemistry* **40**, 422–428.
 - 179 Tetreault M, Cusanovich M, Meyer T, Axelrod H & Okamura MY (2002) Double mutant studies identify electrostatic interactions that are important for docking cytochrome c (2) onto the bacterial reaction center. *Biochemistry* **41**, 5807–5815.
 - 180 Kersten B, Possling A, Blaessing F, Mirgorodskaya E, Gobom J & Seitz H (2004) Protein microarray technology and ultraviolet crosslinking combined with mass

- spectrometry for the analysis of protein–DNA interactions. *Anal Biochem* **331**, 303–313.
- 181 Benjamin DC, Williams DC, Smithgill SJ & Rule GS (1992) Long-range changes in a protein antigen due to antigen–antibody interaction. *Biochemistry* **31**, 9539–9545.
- 182 Song J & Markley JL (2001) NMR chemical shift mapping of the binding site of a protein proteinase. *J Mol Recognit* **14**, 166–171.
- 183 Morrison J, Yang JC, Stewart M & Neuhaus D (2003) Solution NMR study of the interaction between NTF2 and nucleoporin FxFG. *J Mol Biol* **333**, 587–603.
- 184 Foster MP, Wuttke DS, Clemens KR, Jahnke W, Radhakrishnan I, Tennant L, Reymond M, Chung J & Wright PE (1998) Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *J Biomol NMR* **12**, 51–71.
- 185 Ramos A, Kelly G, Hollingworth D, Pastore A & Frenkiel T (2000) Mapping the interfaces of protein–nucleic acid complexes using cross-saturation. *J Am Chem Soc* **122**, 11311–11314.
- 186 Schubert M, Edge RE, Lario P, Cook MA, Strynadka NCJ, Mackie GA & McIntosh LP (2004) Structural characterization of the RNase E S1 domain and identification of its oligonucleotide-binding and dimerization interfaces. *J Mol Biol* **341**, 37–54.
- 187 Fields BA, Goldbaum FA, Ysern X, Poljak RJ & Mariuzza RA (1995) Molecular-basis of antigen mimicry by an anti-idiotope. *Nature* **374**, 739–742.
- 188 Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* **24**, 946–950.
- 189 Merrit EA & Murphy MEP (1994) Raster3D, Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr D* **50**, 869–873.
- 190 Bressanelli S, Stiasny K, Allison SL, Stura EA, Duquerroy S, Lescar J, Heinz FX & Rey FA (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J* **23**, 728–738.