PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Solvated protein–protein docking using Kyte-Doolittle-based water preferences

Panagiotis L. Kastritis,[1] Koen M. Visscher,[1] Aalt D. J. van Dijk,[2] and Alexandre M. J. J. Bonvin[1]*

[1] Bijvoet Center for Biomolecular Research, Faculty of Science, Department of Chemistry, Utrecht University, Padualaan 8, 3584CH, Utrecht, The Netherlands

[2] Applied Bioinformatics, Department of Bioscience, Wageningen UR, Droevendaalsesteeg 1, 6708PB, Wageningen, The Netherlands

ABSTRACT

HADDOCK is one of the few docking programs that can explicitly account for water molecules in the docking process. Its solvated docking protocol starts from hydrated molecules and a fraction of the resulting interfacial waters is subsequently removed in a biased Monte Carlo procedure based on water-mediated contact probabilities. The latter were derived from an analysis of water contact frequencies from high-resolution crystal structures. Here, we introduce a simple water-mediated amino acid–amino acid contact probability scale derived from the Kyte-Doolittle hydrophobicity scale and assess its performance on the largest high-resolution dataset developed to date for solvated docking. Both scales yield high-quality docking results. The novel and simple hydrophobicity scale, which should reflect better the physicochemical principles underlying contact propensities, leads to a performance improvement of around 10% in ranking, cluster quality and water recovery at the interface compared with the statistics-based original solvated docking protocol.

Proteins 2013; 81:510–518.

Key words: Kyte-Doolittle hydrophobicity scale; explicit water; macromolecular docking; structure; protein–protein interactions.

INTRODUCTION

Water is essential to life, being an active constituent in all levels of complexity, from molecules and cells to tissues and organisms: Water not only participates as a solvent, but as an active matrix, since almost all biomacromolecules are inactive in its absence.[1] Hydration determines their structural stability, dynamics, and function *in vitro* and *in vivo*.[2] Particularly for proteins, structure and dynamics of water–protein interactions underline a plethora of molecular phenomena, including protein folding,[3] maintenance of structural integrity,[4] acceleration of enzymatic catalysis,[5] and mediation of molecular recognition.[1] Therefore, it is of great importance to characterize and model the dynamic behavior of a biomolecule-associated waters, the so-called biological waters, at a molecular level.[6] Previous work has already highlighted that water can directly influence the structure and energetics of the interface.[7] As an example, the first crystal structure of a protein–protein complex,[8] that of trypsin with the pancreatic trypsin inhibitor (PTI), already clearly demonstrated the role of water (Fig. 1): The Asp_{189} side-chain of trypsin is in contact with the Lys_{15} side-chain of PTI via water-mediated hydrogen bonds. In contrast, in the structure of trypsin in complex with the homologous inhibitor from soybean (STI), the same water molecule is absent, the salt bridge being formed directly via the bulkier positively charged Arg residue of STI that substituted Lys_{15}.

Water is also essential for transient protein-protein interactions that form the basis of almost every molecular process in the cell.[9] Even though experimental structural characterization of weak binders still remains difficult,[10] understanding the structural details of these interactions should open the route to novel therapeutics through structure-based drug design.[11] Computational protein–protein docking provides a complementary approach towards the characterization of such interactions, especially when some experimental data are available. These can be used either *a posteriori* to select models, or *a priori*, to drive the entire docking procedure.[12] HADDOCK,[13] a data-driven

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Netherlands Organization for Scientific Research (NWO); Grant number: 700.56.442

*Correspondence to: Alexandre M. J. J. Bonvin, Bijvoet Center for Biomolecular Research, Faculty of Science, Chemistry, Utrecht University, Padualaan 8, 3584CH, Utrecht, The Netherlands. E-mail: a.m.j.j.bonvin@uu.nl

Received 13 July 2012; Revised 6 October 2012; Accepted 21 October 2012
Published online 15 November 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24210

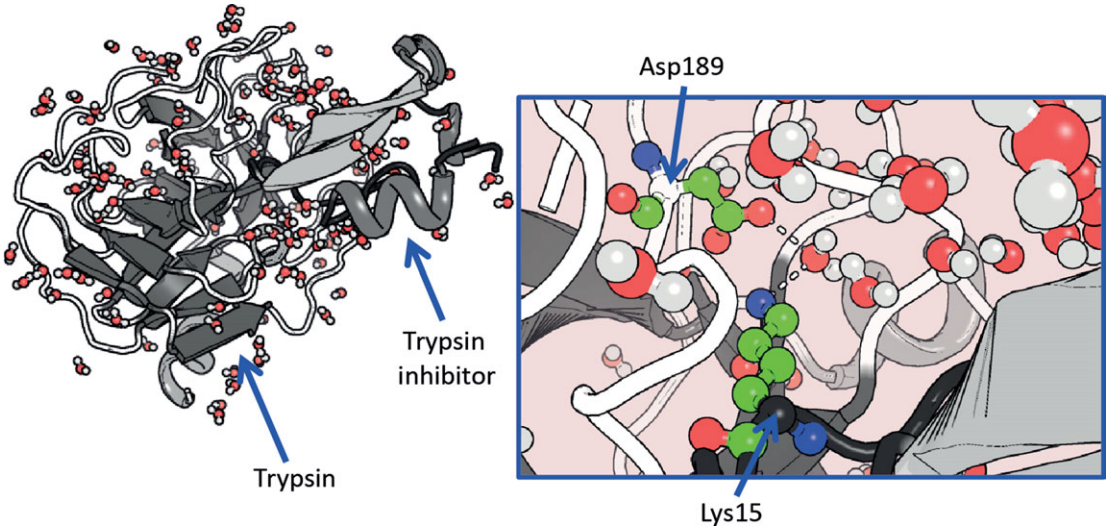

Figure 1

Role of water in protein–protein complexes as illustrated by the water-mediated hydrogen bonds in the Trypsin-PTI complex (PDB entry 2PTC).[8] Figure generated with Pymol (http://www.pymol.org/).

macromolecular docking program, falls into the latter category, incorporating *a priori* experimental information from various sources to guide the modeling process. The HADDOCK protocol consists of a rigid-body docking step followed by semiflexible refinement and final refinement in explicit water.

When it comes to describing water-mediated hydrogen bonds in the interface of macromolecular complexes, HADDOCK is one of the few approaches that allows for explicit solvated docking of proteins by simulating the encounter complex.[14] In that case, the docking is performed from fully solvated proteins. The resulting waters in the generated interface are subsequently removed via a biased Monte Carlo (MC) procedure based on statistical propensities of water-residue contacts in protein interfaces. Finally, an energetic criterion is applied to remove all unfavorable waters in the interface (those with $E_{vdW} + E_{elec} > 0$ kcal/mol). The success of data-driven docking (in terms of number of near-native decoys, scoring and water recovery) has been shown to be influenced by the available information.[14,15] This might, however, also well be influenced by the propensity scale used in the MC water removal process.[14,15]

In this work we investigate the effect of the water-mediated contact propensity scale on the docking performance. We introduce a novel propensity scale based on the Kyte-Doolittle hydrophobicity scale,[16] which is extended to account for water-mediated interface contacts (K-$D_{p\text{-}p}$ scale). In this K-$D_{p\text{-}p}$ scale, both water-vapor transfer free energies and interior–exterior distribution of amino acid side-chains determined by Chothia[17] are used for each amino acid. The resulting water-mediated interface contact probabilities are used during the MC water removal step in the docking process, in the same manner as the previously described statistical scale. The method is tested on a high-resolution solvated docking benchmark consisting of 43 complexes with known bound and unbound partners. Analysis of the performance of solvated docking on this extended benchmark shows that solvated docking does improve protein-protein complex prediction, especially when using the simple K-$D_{p\text{-}p}$ scale.

MATERIALS AND METHODS

Solvated docking protocol

Our solvated docking protocol has been described in detail previously.[15] Briefly, it is composed of four steps:

1. Solvation. Proteins are solvated in a shell of TIP3P water.[18] Waters closer than 4 Å or further away than 8 Å from the protein surface (measured between heavy atoms) are removed and the remaining waters are subjected to a short molecular dynamics optimization in order to allow water to reorient and favorably interact with the protein surface. Waters further away than 5.5 Å from the protein after that stage are removed. This procedure results in a thin and optimized layer of explicit water molecules around each biomolecule.
2. Initial rigid-body docking. Docking starts by rigid body minimization, where each protein with its corresponding solvation shell is treated as one rigid entity. The resulting complexes have thus two possibly overlapping solvation shells. All noninterfacial water

molecules are first removed and the remaining waters together with the protein chains are each treated as separate molecules in a subsequent rigid body energy minimization stage.

3. Monte Carlo-based water removal. Since waters are typically not able to escape the interface during the first rigid-body minimization, further removal is needed. This is achieved in a Monte Carlo approach, removing water until a predefined percentage of the interface remains solvated (25% by default based on the work of Rodier *et al.*[19]). For this, a random water molecule is selected and its closest neighbors on the various protein molecules are identified. The resulting water-mediated interaction pair is assigned a probability to be kept that is derived from the corresponding observed frequencies stored in a database file. This process is repeated until the user-defined cutoff that defines the amount of the interfacial water molecules that must be kept is reached.
4. Energetic criterion. The interaction energy ($E_{vdW} + E_{elec}$) of each water with all protein chains is calculated and all water molecules with positive interaction energy are removed. This process allows potential removal of all waters if unfavorable. The remaining waters and the protein chains are then subjected to a final rigid body energy minimization, with each molecule treated again as a separate rigid body.

The solvated docking protocol described above corresponds to the rigid body docking stage in HADDOCK (it0 step). The subsequent stages of the docking protocol, namely semiflexible refinement in torsion angle space (it1 step) and fully flexible refinement in explicit water, are performed as previously described.[13] The method was originally described by van Dijk and Bonvin[14] and its application to the Barnase-Barstar complex in which water plays a dominant role in the interaction has been illustrated in detail.[15] Overall, the described protocol approximates the general view of protein–protein complex formation: the initial stage mimics the formation of a transient and dynamic encounter complex mainly driven by diffusion and electrostatics; by water removal and refinement the complex then proceeds to form a more stable and tightly packed well-defined form where hydrophobicity plays a more important role.

The solvated docking protocol of HADDOCK is available both via its web-server implementation[20] (http://haddock.science.uu.nl/) and in local mode. In this work, the web-server implementation was used under *guru* access.

Definition of docking parameters and comparison of protocols

Benchmarking was performed with the following HADDOCK settings in order to mimic a realistic docking simulation:

1. unbound conformations of the reactant molecules were used
2. active residues used as restraints to drive the docking procedure were defined as those being in contact in the bound complex using a 5 Å cut-off; 50% of the restraints were randomly removed for each docking trial.
3. passive residues were defined automatically via the web-server, as the neighboring solvent accessible ones.[20]

The various protocols were run with the same number of structures at all stages of the docking (1,000/200/200 for it0, it1 and the water refinement stage, respectively). RMSD-based clustering of solutions was performed with the default 7.5 Å cut-off. The HADDOCK score was used to rank the generated solutions after each stage as previously described.[13–15,20]

Using these parameters, four different docking protocols were evaluated on the benchmark of 43 protein–protein complexes:

1. Unsolvated docking (VAC): Standard HADDOCK protocol – no solvated docking.
2. Solvated docking with the statistical scale (STAT): Solvated docking using the original water-mediated contact propensity scale derived from the Keskin dataset.[21]
3. Solvated docking with the hydrophobicity scale (K-$D_{p\text{-}p}$): Solvated docking using the new water-mediated contact propensity scale derived from the Kyte-Doolittle scale.
4. Solvated docking with a reversed K-$D_{p\text{-}p}$ scale (REV): Solvated docking with an "inverse" water-mediated contact propensity scale derived from reversed values of the normal K-$D_{p\text{-}p}$ scale (K-$D_{p\text{-}p}^{inverse}$). This was done to assess if differences between results derived from STAT and *K-$D_{p\text{-}p}$* are significant. We would expect lower quality docking results with this inverse scale, and testing this allows to further assess the influence of the propensity scale on docking.

Derivation of Kyte-Doolittle-based water preferences (K-$D_{p\text{-}p}$ scale)

According to the Kyte and Doolittle[16] scale, each amino acid has a hydrophobicity value derived by considering both experimentally measured water-vapor transfer free energies for amino acids[22] and observed interior-exterior distribution of amino acid side-chains determined by Chothia.[17]

The probability of a residue to form a hydrogen bond with water, *P(A)*, is assumed independent from the probability of another residue to form another hydrogen bond with the water molecule, *P(B)*. Therefore, if both

occur, $P(AB) = P(A)\,P(B)$. Since we are considering that two residues must form hydrogen bonds to one water molecule, their hydrophilicity is directly related to their ability to form hydrogen bonds with the solvent.

For two amino acids i and j with hydrophobicity values A_i and A_j, a water-mediated contact probability $H_{i,j}$ is calculated as:

$$H_{i,j} = A_i \cdot A_j$$

The $H_{i,j}$ values are subsequently normalized so that:

$$H_{i,j}' \in [0,\ 0.73]$$

where $H_{i,j}'$, denotes the probability of waters to be kept when present between the i and j amino acid residues. The highest probability was set to 0.73 since this corresponds to the maximum observed probability from an analysis of the PDB entries in the Keskin dataset as extensively discussed previously for the statistical scale.[14]

Assessment criteria

Generated models

Standard CAPRI criteria were used to assess the quality of the generated models.[23] Briefly, an acceptable, medium, or high-quality docking prediction (*, **, ***, respectively) are defined by applying the following criteria sequentially:

- Acceptable prediction (*): (i-RMSD $\leq$ 4 Å or l-RMSD $\leq$ 10 Å) and Fnat $\geq$ 0.1
- Good prediction (**): (i-RMSD $\leq$ 2 Å or l-RMSD $\leq$ 5 Å) and Fnat $\geq$ 0.3
- High quality prediction (***): (i-RMSD $\leq$ 1 Å or l-RMSD $\leq$ 1 Å) and Fnat $\geq$ 0.5.

i-RMSD refers to the interface root mean square deviation (RMSD) calculated over the backbone atoms of all residues within 10 Å of the partner molecule, l-RMSD to the ligand RMSD, calculated over the backbone atoms of the ligand (the smallest component) after fitting on the backbone atoms of the receptor, and Fnat corresponds to the fraction of native interatomic contacts calculated using a 5 Å distance cutoff.

Generated clusters

A cluster of acceptable quality denotes a cluster in which at least one of the Top 4 structures in terms of HADDOCK score is below the threshold for an acceptable structure (i-RMSD $\leq$ 4.0 Å).

Water recovery

The water recovery was assessed following the CAPRI criteria set for the analysis of water recovery in Target 47 (see http://www.ebi.ac.uk/msd-srv/capri/round23/). For this, the fraction of water-mediated contacts, f^{w}_{nat}, is calculated by comparison with the known target. Water mediated contacts are defined between residues of the receptor and the ligand that are within 3.5 Å distance of the same water molecule. Five categories are defined depending on the recovery fraction:

- Bad: $f^{w}_{nat} < 0.1$
- Fair: $0.1 \leq f^{w}_{nat} < 0.3$
- Good: $0.3 \leq f^{w}_{nat} < 0.5$
- Excellent: $0.5 \leq f^{w}_{nat} < 0.8$
- outstanding: $f^{w}_{nat} \geq 0.8$

Success rate

For model generation, the success rate is defined as the percentage of benchmark cases for which at least one structure of a corresponding quality has been generated. For cluster quality, the success rate is defined as the percentage of benchmark cases for which at least one cluster of a corresponding quality has been generated in, for example, the Top 2 or Top 1, as ranked accordingly to the HADDOCK score.

RESULTS

The solvated docking benchmark

The new benchmark of 43 complexes includes the 10 original complexes from van Dijk and Bonvin[14] and 33 newly defined protein–protein complexes. Only crystal structures of bound complexes with resolution of 2 Å or better were considered. For details of the complexes compiled in this work, see Supporting Information Table S1. The 43 high-resolution (<2 Å) complexes cover various degrees of conformational change (see Table I) and consist of 6 antibody-antigen complexes, 19 enzyme-substrate or enzyme-inhibitor complexes, and 18 other complexes, including G-proteins or other receptors.

Distinct features of the solvated scales

The original propensity scale using the MC water removal from the interface is based on observed frequencies of water-mediated hydrogen bonds in high-resolution structures [14] [Fig. 2(A), left] and as such is purely a knowledge-based potential. On the other hand, the newly derived propensity scale, the K-$D_{p\text{-}p}$ scale is based on the Kyte-Doolittle hydrophobicity values [Fig. 2(A), right]. It is directly related to both biophysical measurements and interior–exterior distributions of amino acids in proteins. The K-$D_{p\text{-}p}$ scale has also been normalized so that the maximum observed probability of water-mediated hydrogen bond formation is set to 0.73, a value derived out of observations of residue-water-residue contacts.[14] If the

Table I
Solvated Docking Benchmark[a]

Complex type	N	Free energy of binding (kcal mol^{-1})	Cases with conformational change (>2 Å)	Buried surface area (Å^2)	Water-mediated H-bonds[b]
Antibody-antigen	6	11.5–13.6	0	1,525 ± 406	13 ± 7
Enzyme-inhibitor/substrate	19	7.5–18.6	2	1,682 ± 488	11 ± 6
Others (incl.G-proteins, receptors etc)	18	7.0–15.6	3	2,015 ± 827	9 ± 5

[a]See also Supporting Information Table S1 for the full dataset.
[b]Calculated with HBPLUS (http://www.csb.yale.edu/userguides/datamanip/hbplus/hbplus_descrip.html).

maximum probability were to be set to 1.0, this would imply that some amino acid pairs would always be water-bridged, which is not the case experimentally. Furthermore, this would hamper the docking performance (results not shown). Therefore, the K-D$_{p\text{-}p}$ scale can be considered semi-empirical, and the values included are not prone to change, since they were derived directly from a scale used to annotate hydrophobicity of protein molecules for more than 30 years. Overall, the two scales that describe the ability of two residues to form water-mediated hydrogen bonds exhibit similarities in a global manner (r = 0.58 for all types of contacts) [Fig. 2(B,C)], but they do also include significant differences, especially for critical residues, occasionally found in protein–protein

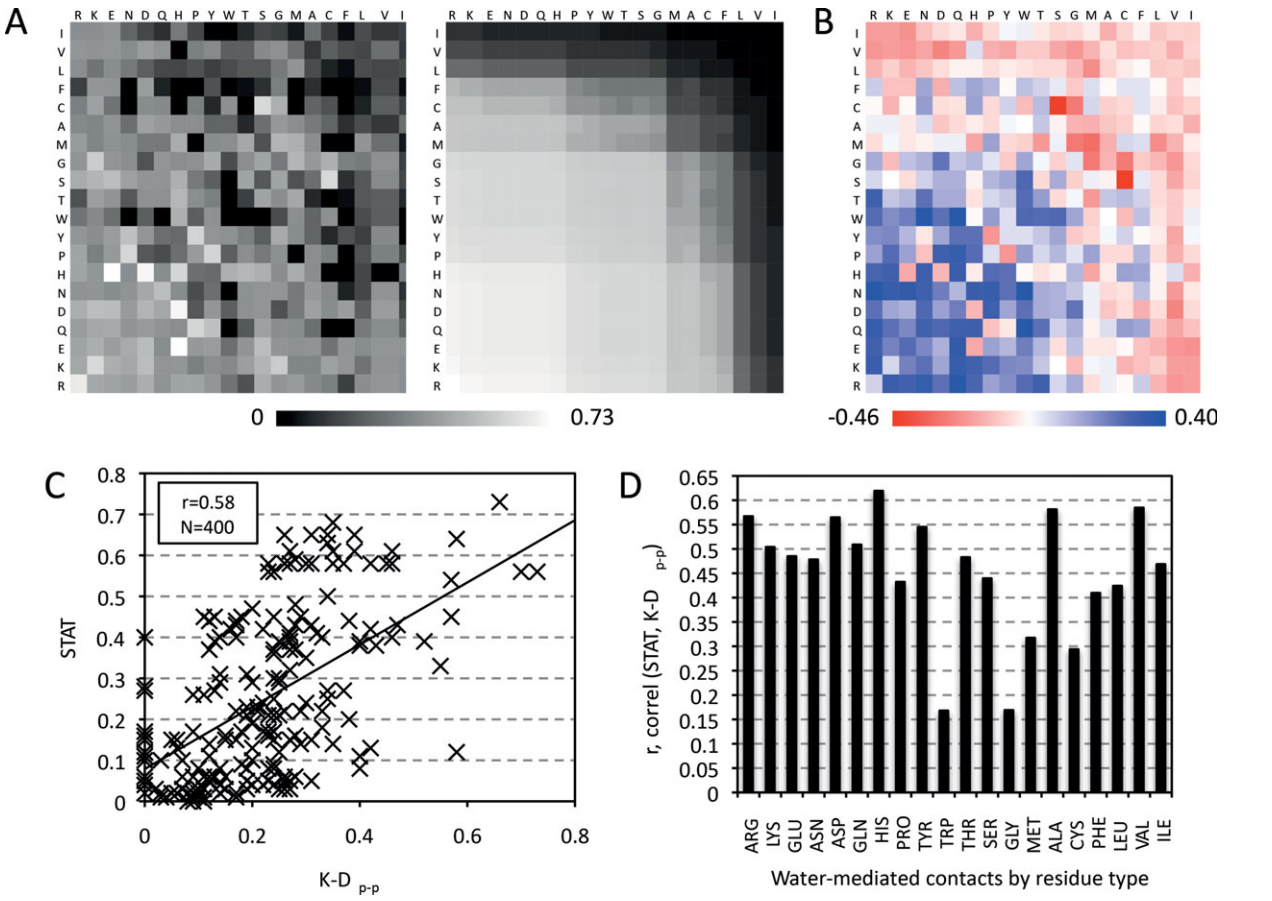

Figure 2

Comparison of solvated docking scales for water-mediated protein–protein interactions. **A**: Propensities of water-mediated contacts following the standard Knowledge-based potential[14] (left) and propensities based on the K-D$_{p\text{-}p}$ scale (right). **B**: Differences in propensities between the knowledge-based and the new K-D$_{p\text{-}p}$ potentials. The color coding goes from Blue to Red, blue indicating high propensities (max = 0.40 for the Asn-Trp and Gln-Trp pairs) in the K-D$_{p\text{-}p}$ scale, whereas red corresponds to lower ones (min = −0.46 for the Ser-Cys pair). **C**: Correlation plot between all pairs of propensities for the statistical (STAT) knowledge-based potential and K-D$_{p\text{-}p}$ potential. **D**: Correlation coefficients for pairs formed for each of the amino acids between the statistical scale and K-D$_{p\text{-}p}$.

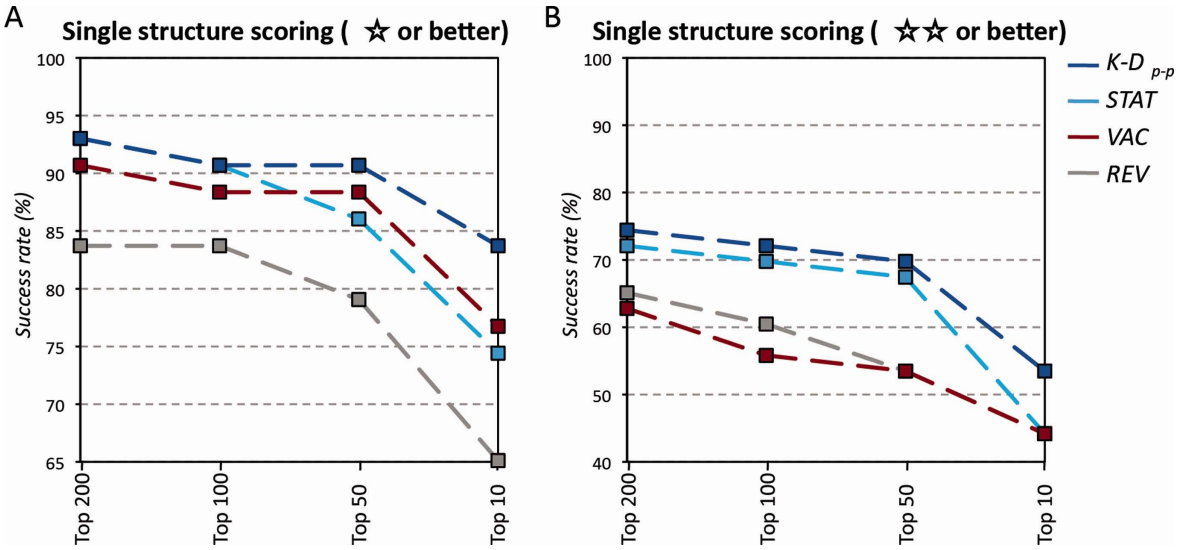

Figure 3

Individual structure generation success rate of the various protocols (43 structures). **A:** Success rate in generating at least an acceptable quality solution or **B:** a medium quality solution in the top 200, 100, 50, and 10, respectively. See Materials and Methods for the quality criteria. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

interfaces, such as Trp ($r < 0.20$), Glu contacts ($r < 0.40$), or important hydrophobic residues (e.g., Phe, $r = 0.41$) [Fig. 2(D)].

Docking results on the benchmark set of 43 complexes

The performance of the solvated docking protocol of HADDOCK, using various water-mediated contact probability scales, was assessed by comparing the results of the four different protocols used: (i) unsolvated docking, (ii) solvated docking with the knowledge-based statistical potential, (iii) solvated docking with the K-$D_{p\text{-}p}$ semi-empirical potential, and (iv) solvated docking with the inverse K-$D_{p\text{-}p}$ potential, the latter as negative control.

Analysis of individual structures

(For details see Supporting Information Table S2). The success rate of the various protocols, defined as the percentage of benchmark cases for which at least one acceptable or better quality model could be generated after water refinement, is shown in Figure 3(A). Considering all 200 models, solvated docking using either the statistical or the Kyte-Doolittle scales performs similarly (93% success rate for acceptable models), only slightly outperforming unsolvated docking (91% success rate). However, this difference increases when considering only the top ranked models where the Kyte-Doolittle scale clearly outperforms all other scales. As also expected, the reverse scale for solvated docking is clearly ranking at the bottom, highlighting that correct solvent structure influences both success rate and ranking in particular.

In Figure 3(B), the same analysis is shown, but for generating and ranking structures of medium quality or better, according to the CAPRI criteria described in Materials and Methods. Again, both solvated protocols perform similarly when considering all 200 models, the K-$D_{p\text{-}p}$ only slightly outperforming the statistical potential already implemented in HADDOCK (75% for the K-$D_{p\text{-}p}$ scale in comparison to 71% for the statistical scale). The performance of unsolvated docking in comparison to solvated docking is however considerably lower. As expected, solvated docking with the reverse scale is once more performing worse compared to the other protocols. When looking at the ranking of the models generated with all scales, the new potential K-$D_{p\text{-}p}$ is always performing substantially better in medium quality structure generation, and especially when looking at the Top 10 medium quality structures (success rate of K-$D_{p\text{-}p}$ is 53%, in comparison to all the other scales that reaches 44% at max).

Cluster Performance

(For details see Supporting Information Table S3). The K-$D_{p\text{-}p}$ scale performs best in generating clusters of acceptable quality overall (Fig. 4). This difference is however not that substantial for the Top cluster compared with unsolvated docking. The statistical scale is performing considerably worse for the Top cluster, and as expected, the reverse scale shows the lowest performance

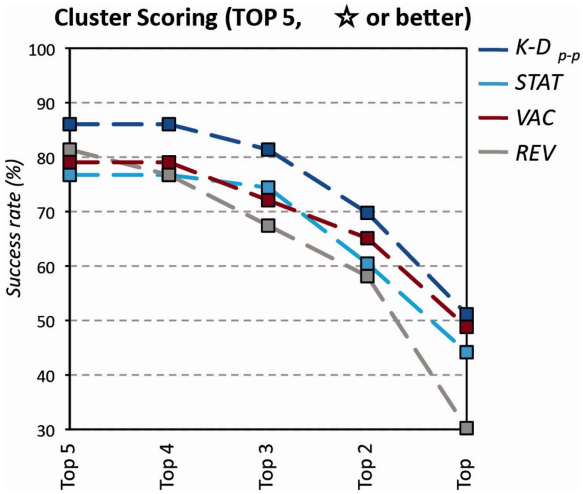

Figure 4

Clustering performance (success rate) expressed in % of the benchmark cases as a function of the number of top clusters considered for all 43 complexes. A cluster is defined acceptable (successful) if one of its top4 members is of acceptable or better quality (see Materials and Methods). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

compared with all other protocols. For the Top 2, all scales generate a cluster of acceptable or better quality in more than 60% of the cases with unsolvated docking and K-$D_{p\text{-}p}$–driven solvated docking demonstrating a success rate above 65%, whereas the knowledge-based potential and the K-$D_{p\text{-}p}^{inverse}$–driven docking show similar rates around 60% (Fig. 4). Overall, the ranking of clusters is always better when using solvated docking with the K-$D_{p\text{-}p}$ scale, independently of the number of clusters selected for comparison.

Water-mediated-contact analysis

(For details see Table S4). The water-mediated contacts were defined using the standard criteria defined by the CAPRI committee for T47 and the performance of water recovery was classified into five categories, depending on the fraction of contacts recovered (see Materials and Methods). When calculating f^{w}_{nat} for all the structures included in the Top cluster according to the HADDOCK score, the newly derived potential K-$D_{p\text{-}p}$ exhibits better water recovery compared with the other scales (knowledge-based and K-$D_{p\text{-}p}^{inverse}$) [Fig. 5(A)]. The performance of the statistical scale is only slightly better than the K-$D_{p\text{-}p}^{inverse}$ scale.

Analysis of the recovery of water-mediated contacts in the Top 4 structures of the best ranked cluster shows a higher recovery of "good" ($0.3 \leq f^{w}_{nat} < 0.5$) contacts for clusters stemming from solvated docking using K-$D_{p\text{-}p}$ [Fig. 5(B)]. We also observe that ~10% of the Top 4 structures generated for all protein–protein complexes exhibit excellent ($0.5 \leq f^{w}_{nat} < 0.8$) recovery of water-mediated contacts, indicating that the HADDOCK ranking within a cluster selects models with better water contact recovery compared with the average recovery within the cluster. Overall, 31% of the clusters have "good" or "excellent" recovery when applying K-$D_{p\text{-}p}$ whereas percentages drop significantly when the other solvated docking scales are applied (22% and 20%, for the knowledge-based potential and K-$D_{p\text{-}p}^{inverse}$, respectively). Further, the number of waters recovered by all solvated docking protocols is related to the number of interfacial waters in the crystal structures tested in this study. For the K-$D_{p\text{-}p}$ protocol in particular, the correlation between the number of crystal waters and the pre-

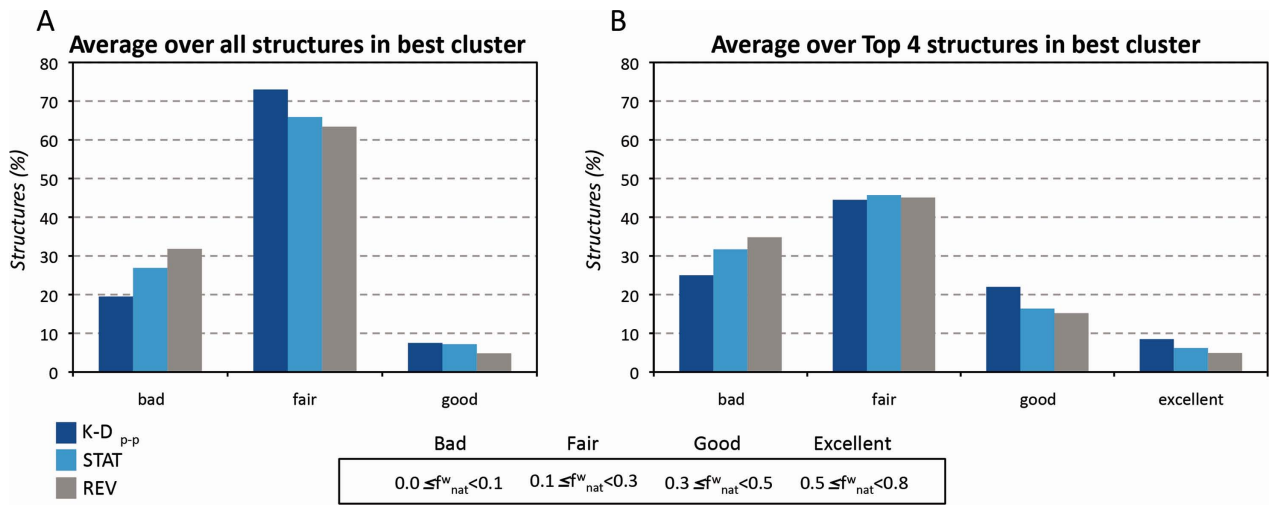

Figure 5

A: Percentage of structures found with the corresponding quality of interfacial water in the Top cluster; Comparison of the three solvated scales; **B:** same as (A), but the analysis was restricted to the Top 4 members of a cluster. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

dicted waters recovered in the Top 4 structures yields $r = 0.48$, with a significant two-tailed P-value of 0.0016. (For all protocols, compare columns 2 and 3 in Supporting Information Table S4).

DISCUSSION

Overall, the new propensity scale, inspired from the Kyte-Doolitle hydrophobicity scale, represents a useful addition to the solvated docking protocol of HADDOCK, achieving better performance compared to the previously published contact statistics-based solvated docking protocol[14] and standard unsolvated docking.[24] The improvement observed by using the new scale K-D$_{p\text{-}p}$ is substantial, resulting in an overall increase of 10% in success rate in terms of both single structure scoring and clustering performance. Although 43 protein–protein complexes remains a limited dataset to use as a benchmark, there are not so many high resolution complexes available for which both free and bound forms are available, a requirement to properly assess the recovery of water-mediated waters in the interface.

Although the two water-mediated contact probability scales are rather similar ($r = 0.56$), the differences noted for several amino acids must be responsible for the improvement in the docking predictions using the Kyte-Doolitle-derived scale, K-D$_{p\text{-}p}$. Several factors might actually limit the derivation of water-mediated contact probabilities from a statistical analysis of high resolution crystal structures. First, some residues might be under- or over-represented when compiling such datasets and consequently the derived contact statistics might not be very accurate. A second related limitation for purely statistical scales is that different water-mediated contact propensities might be derived depending on the data set used,[1,14,25] resulting in (a) tedious comparison of different knowledge-based potentials and (b) change in their derived statistics if reanalyzed using a newer dataset of protein structures. A third and last limitation of a knowledge based potential is that not all water molecules are typically visible in a crystal structures,[26] some of these might actually even be ions (for example it is very difficult to distinguish a Calcium ion from a water). For all these reasons, statistical scales are bound to change as newer complex structures are released.

The implementation of a semi-empirical potential for solvated docking based just on the hydrophobicity of amino acids as originally described by Kyte and Doolittle[16] overcomes these problems. The values in the Kyte and Doolittle scale have been empirically derived from actual biophysical measurements in combination with accepted interior–exterior propensities of residues for proteins. Although it was not optimized for protein–protein interfaces, its simplicity and robustness is directly reflected in the improvement of the docking performance shown here. Several other frequently used hydrophobicity scales have been reported, namely the Eisenberg,[27] Rose,[28] Janin,[29] and Engelman.[30] We observe a high correlation between the chosen Kyte-Doolittle hydrophobicity values and those calculated from the other scales ($0.88 > r > 0.84$, $N = 20$), in contrary to the modest correlation of Kyte-Doolittle with the previously implemented scale in HADDOCK ($r = 0.58$). We therefore expect those scales to perform similarly when implemented in docking. Small differences should, however be expected considering the high, but imperfect relations between the various hydrophobicity scales.

Most work to advance from *in vacuo* to *in solution* docking for protein–protein recognition has concentrated so far on implicit solvation models, used to "preaverage" solvent behavior and thus reduce computationally expensive sampling (reviewed by Chen et al.[31]). Explicit models are often not considered mainly due to the inclusion of higher complexity and dimensionality into the system under study, despite their higher precision and detail.[31] The importance of water in docking has however been recognized for protein–ligand systems[32–34] although its explicit inclusion in protein–ligand docking is still a challenging issue.[35,36] Despite that, explicit water has been shown to significantly improve both docking and scoring of protein–ligand poses, leading to similar improvement rates in prediction performance as the ones demonstrated here for protein–protein complexes.[34,37,38] Along the same lines, explicit treatment of water in macromolecular complexes is getting increased attention in the modeling of macromolecular complexes as demonstrated by a very recent CAPRI round in which interfacial water in a Colicin-Immunity protein complex had to be predicted (Round 23, Target 47, http://www.ebi.ac.uk/msd-srv/capri/round23/).

We believe that the solvated docking benchmark, the simple Kyte-Doolittle water mediated contact potential and the performance of various solvated docking protocols presented in this work should provide the basis for further development and application of explicit solvated docking for macromolecular complexes. Solvated docking with the new K-D$_{p\text{-}p}$ scale will be available in the HADDOCK2.2 distribution. (http://haddock.science.uu.nl/). The propensity scale, together with various docking results statistics, is provided as Supporting Information material.

REFERENCES

1. Levy Y, Onuchic JN. Water mediation in protein folding and molecular recognition. Annu Rev Biophys Biomol Struct 2006;35:389–415.
2. Ball P. Water as an active constituent in cell biology. Chem Rev 2008;108:74–108.
3. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol 2009;16:574–581.
4. Zhang L, Yang Y, Kao YT, Wang L, Zhong D. Protein hydration dynamics and molecular mechanism of coupled water-protein fluctuations. J Am Chem Soc 2009;131:10677–10691.

5. Grossman M, Born B, Heyden M, Tworowski D, Fields GB, Sagi I, Havenith M. Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site. Nat Struct Mol Biol 2011;18:1102–1108.
6. Mentré P. Water in the orchestration of the cell machinery. Some misunderstandings: a short review. J Biol Phys 2012;38:13–26.
7. Janin J. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure 1999;7:R277–R279.
8. Huber R, Kukla D, Bode W, Schwager P, Bartels K, Deisenhofer J, Steigemann W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 A resolution. J Mol Biol 1974;89:73–101.
9. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein-protein interactions: structural, functional, and network properties. Structure 2010;18:1233–1243.
10. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. Cell 2011;144:986–998.
11. Mullard A. Protein-protein interaction inhibitors get into the groove. Nat Rev Drug Discov 2012;11:173–175.
12. van Dijk AD, Boelens R, Bonvin AM. Data-driven docking for the study of biomolecular complexes. FEBS J 2005;272:293–312.
13. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–1737.
14. van Dijk AD, Bonvin AM. Solvated docking: introducing water into the modelling of biomolecular complexes. Bioinformatics 2006;22: 2340–2347.
15. Kastritis PL, van Dijk AD, Bonvin AM. Explicit treatment of water molecules in data-driven protein-protein docking: the solvated HADDOCKing approach. Methods Mol Biol 2012;819:355–374.
16. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
17. Chothia C. The nature of the accessible and buried surfaces in proteins. J Mol Biol 1976;105:1–12.
18. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc 1981;103:335–340.
19. Rodier F, Bahadur RP, Chakrabarti P, Janin J. Hydration of protein-protein interfaces. Proteins 2005;60:36–45.
20. de Vries SJ, van Dijk M, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 2010;5: 883–897.
21. Keskin O, Tsai CJ, Wolfson H, Nussinov R. A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications. Protein Sci 2004;13:1043–1055.
22. Wolfenden RV, Cullis PM, Southgate CC. Water, protein folding, and the genetic code. Science 1979;206:575–577.
23. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2009;78:3073–3084.
24. de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AM. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. Proteins 2007;69:726–733.
25. Papoian GA, Ulander J, Wolynes PG. Role of water mediated interactions in protein-protein recognition landscapes. J Am Chem Soc 2003;125:9170–9178.
26. Levy Y, Onuchic JN. Water and proteins: a love-hate relationship. Proc Natl Acad Sci USA 2004;101:3325–3326.
27. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 1984;179:125–142.
28. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. Science 1985;229:834–838.
29. Janin J. Surface and inside volumes in globular proteins. Nature 1979;277:491–492.
30. Engelman DM, Steitz TA, Goldman A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem 1986;15:321–353.
31. Chen J, Brooks CL, 3rd, Khandogin J. Recent advances in implicit solvent-based methods for biomolecular simulations. Curr Opin Struct Biol 2008;18:140–148.
32. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. Proteins 2006;65:15–26.
33. Lie MA, Thomsen R, Pedersen CN, Schiott B, Christensen MH. Molecular docking with ligand attached water molecules. J Chem Inf Model 2011;51:909–917.
34. de Beer SB, Vermeulen NP, Oostenbrink C. The role of water molecules in computational drug design. Curr Top Med Chem 2010;10:55–66.
35. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. Angew Chem Int Ed Engl 2003;42:2718–2736.
36. Waszkowycz B, Clark DE, Gancia E. Outstanding challenges in protein–ligand docking and structure-based virtual screening. WIREs Comput Mol Sci 2011;1:229–259.
37. Forli S, Olson AJ. A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. J Med Chem 2012;55:623–638.
38. Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, Nissink JW, Taylor RD, Taylor R. Modeling water molecules in protein-ligand docking using GOLD. J Med Chem 2005;48:6504–6515.