This is the post-print version of the following article:

- E. Deplazes, J. Davies, **A.M.J.J. Bonvin**, G.F. King and A.E. Mark. On the Combination of Ambiguous and Unambiguous Data in the Restraint-driven Docking of Flexible Peptides with HADDOCK: The Binding of the Spider Toxin PcTx1 to the Acid Sensing Ion channel (ASIC)1a. *J. Chem. Inf. and Model. 56*, 127-138 (2016). DOI: 10.1021/acs.jcim.5b00529

On the Combination of Ambiguous and Unambiguous Data in the Restraint-driven Docking of Flexible Peptides with HADDOCK: The Binding of the Spider Toxin PcTx1 to the Acid Sensing Ion channel (ASIC) 1a

*Evelyne Deplazes[1,2], Josephine Davies[2], Alexandre M. J. J. Bonvin[3], Glenn F. King[1], *Alan E. Mark[2]

[1]Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia.

[2]School of Chemistry & Molecular Biosciences, The University of Queensland, St. Lucia, QLD 4072, Australia.

[3]Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, 3584 CH Utrecht, the Netherlands.

Keywords: docking, restraint-driven docking, HADDOCK, alanine scanning, peptide binding, acid sensing ion channels, venom peptides, interface residues, structure prediction

List of Abbreviations:

AIR	Ambiguous interaction restraint
ASIC	Acid sensing ion channel
cASIC1	Chicken acid sensing ion channel
ASM	Alanine scanning mutagenesis
HADDOCK	High Ambiguity Driven biomolecular DOCKing
PcTx1	Psalmatoxin-1
RMSD	Root mean square deviation
UDR	Unambiguous distance restraint

Abbreviations used to label the different docking experiments:

AP	Active Peptide
APC	Active Peptide and Channel
C_X	Channel from the PcTx1-cASIC1 co-crystal structure
C_A	Channel from the crystal structure of the apo-form cASIC1a
NR	No removal (of AIRs)
P_X	Peptide from the PcTx1-cASIC1 co-crystal structure
P_N	Consensus NMR structure of PcTx1
P_{20}	Ensemble of 20 models from the NMR structure of PcTx1
UR	Unambiguous Restraints

For example, C_AP_N-1UR-3APC refers to a docking experiment in which the consensus NMR structure of PcTx1 was docked to the channel structure from the PcTx1-cASIC1 co-crystal structure using ambiguous interaction restraints based on 3 active peptide and 3 active channel residues and 1 unambiguous restraint.

ABSTRACT

Peptides that bind to ion channels have attracted much interest as potential lead molecules for the development of new drugs and insecticides. However, the structure determination of large peptide-channel complexes using experimental methods is challenging and structural models are often derived from combining experimental information with restraint-driven docking approaches. Using the complex formed by the venom peptide PcTx1 and acid sensing ion channel (ASIC) 1a as a case study, we have examined the effect of different combinations of restraints and input structures on the statistical likelihood of a) correctly predicting the structure of the binding interface and b) the ability to predict which residues are involved in specific pairwise peptide-channel interactions.

For this, we have analysed over 200'000 water-refined docked structures obtained with various amounts and types of restraints of the peptide-channel complex predicted using the docking program HADDOCK. We found that increasing the number of restraints or even the use of pairwise interaction data did not significantly improve the likelihood of finding a structure within a given accuracy. This suggests that shape complementarity and the force field make a large contribution to the accuracy of the predicted structure. The results also showed that there are large variations in the accuracy of the predicted structure depending on the precise combination of residues used as restraints. Finally, we reflect on the limitations of relying on geometric criteria such as root-mean square deviations to assess the accuracy of docking procedures for lead optimization. We propose that in addition to currently used measures, the likelihood of finding a structure within a given level of accuracy should be also used to evaluate docking methods.

INTRODUCTION

Venoms from arthropods such as spiders, scorpions and centipedes are a rich source of pharmacologically active peptides with diverse properties ranging from analgesic, antiarrhythmic, antimicrobial to antiparasitic effects. Many venom-derived peptides target ion channels showing remarkable potency and subtype selectivity. As a consequence they have attracted much interest as potential lead molecules for pharmaceutical development.[1, 2] For lead optimisation, a detailed understanding of the interactions between the peptide and the ion channel is required. In recent years, the structures of several ion channels and some peptide-channel complexes have been solved experimentally in atomic detail, providing information on the precise manner in which certain peptides interact with specific ion channels. However, as ion channels are large integral membrane proteins, the routine use of high-resolution structural techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy to analyse alternative complexes is impractical. Instead, information on the nature of the interactions that govern the activity of the peptide is often inferred by combining experimental data from methods such as alanine scanning mutagenesis (ASM) or thermodynamic mutant cycles. This data is then used together with docking approaches to derive a structural model of the peptide-channel complex.[3-9] The question is: how reliable are such models and what is the most effective use of the available experimental data?

There have been many studies demonstrating the ability of various docking approaches to reproduce the structures of complexes solved experimentally. Indeed, over the last decade, significant progress has been made in improving the accuracy of docking methods.[10-16] Nevertheless, predictions involving large flexible ligands such as peptides remain challenging due to the complexity of the binding interface and the large number of potential interactions. It is for this reason that experimental information regarding the binding interface is often incorporated as restraints during the docking process in order to improve the accuracy of the predicted structure. Potential sources of data for restraints include ASM, NMR titrations, crosslinking experiments, and bioinformatics predictions.[13, 17-22] Of these methods, data from ASM of the peptide is perhaps the most readily obtainable.[23, 24] ASM can be used to identify residues that are critical for binding and thus likely to form part of the complex interface. Similarly, mutations of the channel can be used to identify residues within the peptide binding pocket. Such data is, however, is ambiguous in the sense it is not possible to identify which pairs of residues in the peptide and protein interact. To unambiguously identify pairwise interactions between the peptide and protein, function-recovery mutations can be used in which complementary mutations (e.g. charge reversal mutations) are made on both binding partners. This is considerably more resource intensive and technically challenging.

The question is to what extend does the incorporation of specific pairwise interactions improve the reliability with which the interactions between a protein and a peptide ligand can be predicted?

For example, Saez *et. al.* [25] used a restraint-driven docking approach to predict the manner in which the spider venom peptide PcTx1 (Figure 1a) binds to the acid sensing ion channel (ASIC) 1a (Figure 1b). ASICs are cation-selective, proton-gated ion channels associated with nociception and other physiological processes and they are a promising target for the treatment of pain as well as a range of neurological disorders.[26-30] PcTx1 is a 40-residue peptide with a inhibitory cysteine knot motif isolated from the venom of the tarantula *Psalmopoeus cambridgei*. It is the most potent and selective blocker of ASIC1 known.[31] ASM of PcTx1 revealed that residues Trp24, Arg26 and Arg27 are critical for inhibition of rat ASIC1a.[25] This data, in combination with an approximation of the binding site on the channel [32] and a homology model of rat ASIC1a (based on a 1.9 Å crystal structure of the apo form of chicken ASIC1), was used to perform restraint-driven docking to predict the binding mode of PcTx1.[25] Subsequently, the co-crystal structure of PcTx1 in complex with chicken ASIC1 (cASIC1) was published.[33, 34] This confirmed the predicted location of the binding site and the fact that the structure of the binding site was largely unaffected by the presence of the peptide. Indeed, the root mean squared deviation (RMSD) between all residues within 10 Å of the peptide in the complex and those in the apo structure was < 1 Å. However, the docked structure failed to correctly predict the orientation of the peptide in the binding pocket (Figure 1c). The RMSD between the peptide in the top-ranked complex obtained using the docking program HADDOCK and that in the crystal structure when fitting the two complexes based on the receptor was > 6 Å. While this would still fall in a range considered acceptable in some community assessments of docking approaches such as CAPRI: Critical Assessment of PRediction of Interactions initiative [21], the model of Saez *et. al.* was not of sufficient quality to be useful for rational drug design. Reasons why the docking procedure could have failed include: (i) docking was performed using binding data from rat ASIC1a and a homology model of the rat ASIC1a rather than the structure of apo chicken ASIC1 directly; (ii) there was insufficient information to uniquely define the structure of the complex by docking; for example only the general location of the binding pocket was known; (iii) the accuracy of the model was limited by the nature of the docking procedure used or the criteria to judge success; or (iv) the correct binding pose was included in the ensemble of possible structures but was not the highest ranked solution. Other studies that used different docking approaches on the same complex also failed to predict the correct binding mode of PcTx1. [35, 36]

In docking approaches, success is often expressed in terms of whether an appropriate solution can be found within a set of possible solutions. That is, whether one or more of the structures generated is considered sufficiently similar to that determined experimentally.[13, 21] Similarity is often assessed based on geometric criteria such as the RMSD of the peptide and surrounding residues. However, from the perspective of an experimentalist aiming to understand the nature of the interactions within a given complex for which no experimental structure is available the primary questions are: (i) given a particular set of restraints what is the likelihood that the structure(s) produced by a given docking simulation can be used to predict the peptide and channel residues at the binding interface and their pairwise interactions; (ii) to what extent can the reliability of the structures be improved by the incorporation of additional experimental data; and (iii) what is the most effective way a given set of experimental data can be used.

Here we address these questions using the PcTx1-cASIC1 complex as a case study. We compare the effect of different types of restraints on the reliability of the structure of the complex predicted by restraint-driven docking as well as the extent to which the predicted structures can be used to identify pairwise interactions between the peptide and the channel.

EXPERIMENTAL METHODS

All docking runs were performed using the HADDOCK (High Ambiguity Driven biomolecular DOCKing) webserver (version 2.1).[37] HADDOCK combines force field approaches with restraint driven docking and is widely used for the prediction of biomolecular complexes.[19] HADDOCK is designed for use with a range of experimental data including distance and orientation information from NMR as well as data from cross-linking studies or site-directed mutagenesis. Experimental data is translated into either a collection of ambiguous interaction restraints (AIRs) or a number of specific, highly trusted, unambiguous distance restraints (UDRs) that are used to guide the docking process. An AIR is a restraint between a specific residue (or atom) on one molecule (A) and a set of potential interaction partners (residues or atoms) on another molecule (B). The residues at the interfaces are referred to as “active” and “passive”: Active residues are selected by the user and are normally residues that have been shown experimentally to be associated with binding. For example, residues that led to a significant change in binding affinity when mutated to alanine. The corresponding set of passive residues is made up of the potential binding partners for these active residues. Passive residues may either be provided by the user or selected automatically by HADDOCK. AIRs are used to enforce that each active residue is in close proximity (within a given threshold distance) to one or more of the active+passive residues on the partner molecule. For

AIRs, HADDOCK defines this effective distance based on a weighted sum over the distances between a given active residue on one molecule and all active+passive residues on the other. The effective distance for an active residue on A is given by:[17, 19]

$$d_{iAB}^{eff} = \left(\sum_{m_{iA}=1}^{N_{A\,atom}} \sum_{k=1}^{N_{resB}} \sum_{m_{iB}=1}^{N_{B\,atom}} \frac{1}{d_{miA\,nkB}^{6}} \right)^{\left(-\frac{1}{6}\right)} \qquad (1)$$

where $N_{A\ atom}$ refers to all atoms in residue *i* in molecule A, N_{resB} indicates the active and passive residues in molecule B, $N_{B\ atom}$ refers to all atoms of residue *k* in molecule B, and *d* is the distance between atoms m_{iA} and n_{kB}. The effective distance for each active residue in binding partner B is calculated in an equivalent manner. By default, the upper limit of the effective distance is 2.0 Å. This was used for all restraints in all docking simulations. Note the effective distance is always shorter than the shortest distance entering the sum. As many atom-atom distances contribute to the effective distance between residues, an AIR is typically satisfied if an active residue comes within 3–5 Å of an active or passive residue on the other binding partner.[17, 19] In contrast, a UDR is a restraint between a specific pair of residues or atoms. UDRs are used to enforce a given distance between a pair of residues or atoms known (or assumed) to interact in a pairwise manner. For UDRs, the effective distance between the two residues is calculated using all atoms in the residue pair as given in Eq. 1. The upper limit for the UDRs was also 2.0 Å. For a detailed overview of other types of restraints available in HADDOCK see refs.[17, 19, 38]

Overview of the docking runs

An overview of the docking runs is given in Figure 2. Four sets of docking simulations were performed using different combinations of restraints. Each set of docking simulations was further divided into a series of independent trials. In these trials the number of AIRs or UDRs was varied, as was the input structure of the peptide and protein. Each trial consisted of 20 independent docking runs involving different combinations of residues selected at random. In total, over 1200 docking runs were performed and analysed. The residues used to form AIRs and UDRs were based on the 8 peptide and 24 channel residues that were observed to form peptide-channel contacts within the PcTx1-cASIC1 crystal structure interface. [34] Residues for AIRs were randomly selected from these peptide and channel residues or in the case of the UDRs from the corresponding 37 pairwise peptide-channel contacts. These are listed in Table S1 and S2 in the supporting information. Note, during the selection of residue pairs duplicates were eliminated. However, a single residue might still be involved in more than one pair.

The first set of docking simulations corresponded to the case in which specific information is available for the peptide but only the general location of the channel binding site in known. In the individual trials, 3, 4 or 5 peptide residues were selected to be active. These active peptide residues were paired with a set of 31 passive residues on the channel. The latter correspond to all solvent accessible residues within 10 Å of the geometric centre of the binding pocket. These docking simulations are labelled AP for "Active Peptide". The second set of experiments represented a situation in which binding residues have been experimentally identified on both the channel and peptide. In the individual trials, 3, 4 or 5 residues on the peptide as well as an equal number of residues on the channel were considered to be active. The set of passive residues on both the peptide and the channel were assigned automatically by HADDOCK using the default settings. These experiments are labelled APC for "Active Peptide and Channel". Note that, during the docking process, HADDOCK makes multiple attempts to generate a structure of the peptide-protein complex that fulfils the restraints. By default, for each attempt, HADDOCK excludes 50 % of the AIRs. This means each model in the final set of structures produced by HADDOCK will have been calculated using a different subset of the restraints provided by the user. This is to account for possible errors in the AIR data provided. To evaluate the effect of the removal of restraints, the 20 runs in the AP and APC trials were repeated using the same set of restraints as described above but without the removal of any restraints. These runs are labelled NR for "no removal'. The third set of experiments was carried out using unambiguous data only. The individual trials used 1, 2, 3, 4 or 5 pairs of residues that form close contacts across the peptide-channel interface in the PcTx1-cASIC1 crystal structure as UDRs. These experiments are labelled UR for "Unambiguous Restraints". The fourth and final set involved a combination of ambiguous and unambiguous restraints. Specifically, 1 or 2 UDRs were added to a given set of APC restraints as described above.

Each set of docking simulations was also performed using different combinations of input structures. Four alternative combinations of the peptide (P) and channel (C) were examined. The first consisted of the structures of the channel and the peptide extracted from the co-crystal structure (PDB 3S3X).[34] In this case there is an exact conformational match between the structure of the peptide and the binding pocket in the channel. These runs are labelled C_XP_X where the subscript X refers to X-ray. The second combination consisted of the channel extracted from the co-crystal structure and the consensus PcTx1 NMR structure (model 1, PDB 2KNI [25]). These runs are labelled C_XP_N for "X-ray channel and NMR peptide". The third combination, labelled as C_XP_{N20}, consisted of the channel extracted from the co-crystal structure and 20 models from the NMR structure of PcTx1. The fourth combined the crystal structure of the channel in the apo form (crystallised in the

absence of any ligand, PDB 2QTS)[39] with the consensus PcTx1 NMR structure.[39] These runs are denoted C_AP_N for "Apo channel and NMR peptide".

Structure preparation

For the C_XP_X experiments the structures of the channel and the peptide from the crystal structure of the PcTx1-cASIC1 complex (PDB 3S3X)[34] were used. The three subunits forming the channel (chains A, B and C) and one toxin (chain E) were extracted while crystallographic water, glucosamine molecules and ions were removed. In the 3S3X structure a short sequence of 5 residues in the extracellular loop of the channel is missing. These residues were modelled based on the equivalent section from an alternative PcTx1-cASIC1 crystal structure (PDB 4FZ1).[33] These residues are not involved in the binding of PcTx1. For the C_XP_N and C_XP_{N20} docking experiments the same channel structure as in the C_XP_X experiments was used while the peptide structure was taken from the PcTx1 NMR structure PDB 2KNI.[25] For the C_XP_N experiments the consensus PcTx1 NMR structure (model 1) was used while for C_XP_{N20} experiments the first 20 models were used. Finally, for the C_AP_N experiments the channel structure from the crystal structure of cASIC1a in the apo form (PDB 2QTS [39]) was used. In all cases, missing atoms were automatically added by HADDOCK when required.

Analysis of docking runs

The docking process in HADDOCK consists of three stages. First a rigid-body search is performed. By default this generates 1000 structures. Second, the 200 best structures are selected based on the scoring function that includes a combination of van der Waals, electrostatic, empirical desolvation and restraint energy terms together with a buried surface area term. These 200 structures are then used for a semi-flexible refinement in torsion angle space. Third, the 200 structures are refined in explicit water. During the refinement, both side chain and backbone of interface residues are progressively allowed to move. Each of the final 200 water-refined structures is then assigned a HADDOCK score, which is used to rank the structures. HADDOCK also clusters the structures based on RMSD. Each cluster is given a HADDOCK score that is the average of the four best scoring structures in that cluster. The clusters are numbered with respect to size. The 'top-ranked' cluster, however, refers to the cluster with the best HADDOCK score.

In this study, different measures were used to assess the utility of the docked structures. The first was the Root Mean Square Positional Deviation of residues in the binding interface (i-RMSD) between a given docked structure and the PcTx1-cASIC1 co-crystal structure (PDB 3S3X).[34] The

interface was defined as all peptide residues found within 10 Å of any channel residue and all channel residues within 10 Å of any peptide residue in the 3S3X structure. The i-RMSD provides a measure of the extent to which the structure of the interface is predicted correctly. From an engineering perspective, the ability to predict specific interactions between the peptide and channel is equally important. This was assessed in several ways. The first was in terms of the fraction of peptide and channel residues correctly predicted to lie in the peptide-channel interface. The fraction of native peptide interface residues (F_{nat}-PIR) and the fraction of native channel interface residues (F_{nat}-CIR) were calculated using an analogous approach to that used in the definition of i-RSMD except for that fact that a cut-off of 5 Å was used. The interfacial residues in the co-crystal and the predicted complex were determined separately and then compared. In addition, the fraction of pairwise peptide-channel contacts predicted correctly i.e. the 'fraction of native contacts' (F_{nat}) was also determined. Two residues were considered to be in contact if any atom of the two residues were within 5 Å.

The complete set of 200 water-refined structures produced by HADDOCK was compared to the PcTx1-cASIC1 co-crystal structure (PDB 3S3X). However, for simplicity the results presented are based on a single representative structure for each run. Four alternate ways of selecting a single structure for analysis were examined: (i) the top-ranked structure, which is the structure with the lowest HADDOCK score; (ii) the most common structure amongst the 200 water-refined structures as determined using the clustering algorithm described by Daura *et. al.*[40] and an all-atom RMSD cut-off of 1.5 Å; (iii) a structure selected at random from the 200 water-refined structures; and (iv) a single structure selected at random from the four best scoring structures from the top-ranked cluster as provided by HADDOCK.

RESULTS

Peptide data only (AP experiments)

The first set of docking simulations corresponds to where specific information is available for the peptide but only the general location of the binding site is known. Twelve different trials were performed in which 3, 4 or 5 active peptide residues selected at random were combined with a set of passive residues that approximated the binding pocket and different combinations of input structures for the channel and the peptide were used. Each trial consisted of 20 independent docking runs. A summary of the results is shown in Table 1. The results are presented as the percentage of runs in which the i-RMSD between the top-ranked structure (i.e. the top-ranking structure according to the HADDOCK scoring function) and the crystal structure is less than a given cut-off. For ease

of comparison the results are separated into three arbitrary groups (0–35 %, 36–70 % and 71–100 %).

In the first three trials in Table 1 the structure of both the channel and the peptide were extracted from the 3S3X co-crystal structure (C_XP_X-trials). As expected, these trials show the highest probability of finding a structure within a given cut-off of the crystal structure. They are also the only trials that yielded any solutions with an i-RMSD < 2 Å (considering all 200 structures produced by HADDOCK). The probability is insensitive to the number of restraints suggesting that the ability to find appropriate solutions is primarily due to shape complementarity. The next three trials used the channel from the co-crystal structure together with the consensus NMR structure of PcTx1 (C_XP_N-trials). Here there was only a 30 % chance of predicting the structure to within 5 Å i-RMSD. The next three trials used the apo-crystal structure of the channel together with the consensus NMR structure of PcTx1 (C_AP_N-trials). To an experimentalist this is perhaps the most relevant scenario. Again, only 30 % of the structures were within 5 Å of the co-crystal structure. By default, HADDOCK removes 50 % of AIRs during any docking run to account for potential errors in the set of active residues. In this case study, however, the active residues are based on the available co-crystal structure and are thus known to form part of the binding surface. Thus, the automated removal option was disabled in the final set of trials. Comparing the results using the default settings (C_AP_N - AP) with those where all restraints were used (C_AP_N - AP_{NR}) shows that forcing all restraints to be used results in only a small improvement in the probability of finding a correct solution. This suggests that, in this system, shape complementarity and the force field are dominant. The importance of the precise shape of the channel is also evident from the fact that using the apo-structure in which the channel is slightly more open than the co-crystal form leads to better solutions (C_XP_N vs C_AP_N). Comparing the results within the same trial for runs with increasing number of restraints shows that, independent of the structure used, the success rate is mostly insensitive to the number of restraints.

Overall, the probability that the top-ranked structure in any given run has an i-RMSD of less than 3 Å or 4 Å when using mutation data on the peptide only is in the order of 10–15 % and 20–30 %, respectively. Figure 3a shows the top-ranked structures from the 20 runs of the C_AP_N-0UR-4AP trial overlaid on the co-crystal structure. For all runs, the same number of restraints was used and the only difference between the runs was the specific channel and peptide residues involved in the restraints. As can be seen in Fig. 3a, the orientation of the peptide varies significantly. The i-RMSD values range from 2.7 Å to 11.1 Å. This demonstrates that the quality of the prediction depends heavily on the specific combination of restraints. However, no single individual active residue led

to a significantly higher proportion of successful runs. Furthermore, only 6 of the 20 runs in the C_AP_N-0UR-4AP trial produced a top-ranked structure with an i-RMSD < 4 Å. Figure 3b shows an overlay of the top-ranked structure from these 6 runs in comparison to the co-crystal structure. As a reference, structures with i-RMSD values from 1 to 6 Å are provided as supporting information (Figure S1). In the case of the PcTx1-cASIC1 complex, an i-RMSD < 4 Å is sufficient for the peptide to adopt the correct overall position in the binding pocket. Nevertheless, even amongst structures with an i-RMSD < 4 Å there is still some variation in the precise binding mode that could affect the interpretation of experimental data. For example, the fraction of correctly predicted peptide residues, F_{nat}-PIR, is on average 80 % while the fraction of correctly predicted channel residues, F_{nat}-CIR, is on average 60 %. The fraction of correctly predicted pairwise peptide-channel contacts, F_{nat}, varies from 40 to 49%. Note, i-RMSD, F_{nat}-CIR, F_{nat}-CIR and F_{nat} are clearly correlated. A structure with low i-RMSD shows higher values for F_{nat}-PIR, F_{nat}-CIR and F_{nat}. Considering all 20 structures shown in Figure 3a, the minimum value is 60 % for F_{nat}-PIR, 40 % for F_{nat}-CIR and 10 % for F_{nat}.

Peptide and channel data (APC runs)

In the second set of docking simulations, residues on both the peptide and the channel were treated as active. A total of 15 trials were carried out in which 3, 4 or 5 peptides residues were combined with 3, 4 or 5 channel residues. As before, each trial consisted of 20 independent runs. The results for these trials are given in Table 2. In the first three trials the structure of both the channel and peptide from the 3S3X co-crystal structure was used (C_XP_X-trials). Using data from both the peptide and the channel resulted in a significantly higher probability of finding a structure within a given cut-off than the equivalent AP trials shown in Table 1. Again, this was the only case where structures with an i-RMSD < 2 Å were observed. Improvements were also seen in the C_XP_N-trials, in which the channel structure from the co-crystal structure was combined with the consensus NMR structure of PcTx1. Both the C_XP_X-trials and the C_XP_N-trials were relatively insensitive to the number of restraints, again underlining the importance of shape complementarity. To investigate this further, three trials were performed using the crystal structure of the channel with a set of 20 alternative NMR models of PcTx1 (2KN1) (C_XP_{N20}-trials). Using multiple peptide conformations implicitly incorporates a degree of backbone flexibility. The chance of finding a structure with an i-RMSD of < 4 Å increased from 10–20 % using the single consensus NMR structure to 40–50 % using an ensemble of peptide conformations.

In the final two sets of trials of the APC experiments, the apo-crystal structure of the channel was combined with the consensus NMR structure. These two trials were performed with and without the

automatic removal of restraints (C_AP_N-APC and C_AP_N-APC_{NR}, respectively). As in the AP experiments, this option had only a minor effect on the results. A small increase in the chance of finding a correct structure was observed when using 4 or 5 active residues. Overall, comparing the APC results in Table 2 to the AP results in Table 1, the use of active residues on both the peptide and channel doubled the chance of the top-ranked structure being within 3 Å of the co-crystal structure. For example, in the 4AP trials, only 6 of the 20 top-scored structures showed an i-RMSD < 4 Å while in the 4APC trials this increased to 12. For these 12 structures, the average F_{nat}-PIR and F_{nat}-CIR were 78 % and 60 %, respectively, and F_{nat} ranged from 28 % to 51 %. These values are essentially identical to the AP values.

Inclusion of pairwise data (UR and UR-APC experiments)

In the experiments described above the information about which residues interacted in the interface was ambiguous. If information on pairwise interactions is known experimentally, unambiguous distance restraints can be used to guide the docking process. To assess the utility of such additional information, 11 trials with different combinations of UDRs and AIRs were performed. In the first five trials the only restraints were 1–5 UDRs (C_AP_N-UR). In the remaining 6 trials, 1 or 2 UDRs were added to the set of AIRs used in the APC experiments (C_AP_N-1UR-APC and C_AP_N-2UR-APC). All trials were performed using the apo-channel structure and the consensus NMR structure of PcTx1. The results of these trials are shown in Table 3.

As can be seen from Table 3, the probability of finding a structure with a given i-RMSD value using 1 UDR is similar to that obtained using peptide-data only. This again suggests that the docking is dominated by shape complementarity and the force field. The quality of the predictions increased for 2 and 3 UDRs, but was insensitive to additional restraints. Overall, the chances of finding a structure within 3 or 4 Å of the co-crystal structure when using 3, 4 or 5 UDRs ranges between 30 and 65 %, very similar to the results from 4APC and 5APC trials. Adding 1 or 2 UDRs to the set of AIRs used in the C_AP_N-APC trials resulted in only a modest improvement in predictive power. Also, incorporating UDRs did not result in a significant improvement in F_{nat}-PIR, F_{nat}-CIR or F_{nat}.

Alternative structure selection methods

For all results presented so far, the analysis was based solely on the top-ranked structure from each run. In addition to the top-ranked structure (the structure with the lowest HADDOCK score) a series of alternative approaches for determining the best structure were compared. These included: (i) the most common structure amongst the 200 water-refined structures (based on the clustering algorithm

of Daura *et. al.* [40]); (ii) a structure selected at random from the 200 water-refined structures; (iii) a structure selected at random from the top-ranked cluster. The results for these four approaches from the C_AP_N-APC trials are shown in Table 4. Similar trends were observed in all the other trials examined (data not shown). Selecting either the top-ranked structure or the most common structure gave comparable results. Indeed, in many cases similar structures were identified using the two approaches. HADDOCK standardly provides the top four ranked structures for each cluster. Selecting one of these four structures from the top-ranked cluster yielded results comparable to that of the top-ranked or the most common structure. Simply selecting one of the 200 water-refined structures at random resulted in a drop of 50 % in the probability of finding a structure below 4 Å.

DISCUSSION

The primary aim of this study was to understand how alternative combinations of restraints, as might be derived from mutagenesis studies, affect the utility and reproducibility of structural models generated using restraint-driven docking. The specific system considered in this study was the binding of the venom peptide PcTx1 to ASIC1. For this system, both the apo-structure of the channel and the structure of the peptide-channel complex have been solved crystallographically.[33, 34, 39] In addition, an ensemble of high-quality NMR structures for the peptide is available.[25] The PcTx1-ASIC1 system has also been the focus of previous docking studies performed before the co-crystal structure was available. [25, 35, 36]

As expected, if the structures of the peptide and channel were taken directly from the co-crystal structure it was possible to reproduce the experimental binding mode with high accuracy. When data on just the peptide or on both the peptide and channel was used, a significant number of solutions, including many top-ranking solutions, with an i-RMSD < 2 Å were found. However, taking a more realistic scenario in which the structure of the apo-channel, the structure of the peptide in solution, and the general location of the binding site are known, and peptide residues involved in binding were experientially determined (e.g. from ASM), the probability of finding a top-ranking structure with an i-RMSD < 4 Å in this system was in the order of 30 %. No solutions with i-RMSD < 2 Å were found. Given that the structure of the PcTx1-ASIC1 complex proposed by Saez *et. al.*[25] was predicted using peptide-only data and a homology model of rat ASIC1a, the accuracy of the solution obtained was better than might be expected based on this work.

If in addition to data on the peptide it is known precisely which residues on the channel are involved in binding, the probability of finding top-ranking structures with an i-RMSD < 4 Å

increased to around 60 %. Again, no solutions with i-RMSD < 2 Å were found. For many applications, a structure with an i-RMSD of between 2 and 4 Å may well be sufficient. Indeed in the CAPRI community assessment of docked structures a solution with an i-RMSD of ≤ 1 Å is considered to highly accurate, ≥ 1 Å and ≤ 2 Å to be of medium accuracy and ≤ 4 Å to be acceptable.[13] However, there are a number of important caveats. First, the results in this study are based on highly favourable scenarios using ideal data. Second, there were large variations between runs in which the same number of restraints was used but the specific residues varied. For example, it was not possible to identify a single specific residue whose inclusion led to a more accurate prediction. Third, the accuracy of the prediction depended strongly on the input structure of the channel and the peptide. The structure of the binding site in the apo form is essentially identical to that in the complex (heavy atom RMSD < 1 Å). Despite this, significant differences in the accuracy of the docked structure were observed, when using the same consensus NMR structure. Against this, the selection and analysis of the docking model is naïve in the sense only a single model is considered (e.g. the top ranked solution). This approach was taken in order to mimic the approach of a casual or novice user. In practice, when determining specific pairwise interactions using HADDOCK the recommended approach would be to consider multiple solutions, ideally after clustering, and consider all clusters that have comparable scores within their standard deviations (which might or might not give a unambiguous answer). In addition, the backbone RMSD between the consensus NMR structure of the peptide extracted from the co-crystal is 2.9 Å. This is a significant difference. For comparison, in benchmark set of protein complexes widely use to assess the performance of docking programs cases in which the backbone RMSD between the bound and free forms is > 2.5 Å are considered to be challenging.[41] The major differences between the structures are in the β-hairpin loop that contain the majority of the residues known to form interactions with the channel. The sensitivity of the docking procedure to the precise combination of structures and restraints is clearly illustrated in the case of the AP runs involving the co-crystal structures of the peptide and ligand. In approximately half of the runs the i-RMSD of the top ranked structure was < 2 Å while approximately a third of the runs failed to find a reasonable solution (i-RMSD of the top ranked structure < 6 Å). In this case, given a fixed set of restraints, an experimentalist would have ~50 % chance of the structure obtained being useful. Interestingly, the probability of finding solutions below a given i-RMSD value when using the consensus NMR structure of the peptide was significantly higher (almost double) to that using the apo-structure for the channel as opposed to the co-crystal structure. The probability was improved further when using an ensemble of 20 NMR models of PcTx1. This suggests that in this system the apo-structure may be slightly more open than the co-crystal and can thus accommodate alternative conformations of the peptide. The bundle of NMR structures was very tight. The deviation between any pair of NMR

structures (backbone RMSD < 1.5 Å) was much less that the deviation between any of the NMR structures and the co-crystal structure (backbone RMSD between 2.8 Å to 3.1 Å). This suggests that small differences in the structure of the peptide can lead to large differences in the accuracy of the docked complex. An overlay of the binding sites in the apo and holo forms of the channel as well as an overlay of the co-crystal structure and the bundle of NMR structure is provided as supporting information (Figure S2). Note that the trends observed above were largely independent of the number of restraints.

The results described above were obtained using ambiguous data (AIRs). However, equivalent results were obtained using a similar number of unambiguous restraints (UDRs). Combining unambiguous and ambiguous data resulted only in a modest improvement over the use of ambiguous data alone. This might seem surprising. Intuitively one would expect that 3 pairwise contacts would be all that is required to uniquely define the binding pose. This certainly would be the case if both the ligand and channel were rigid, less if the shape complementarity was perfect. However, the shape complementarity is not perfect. In addition, the side chains of residues involved in binding are flexible. As a consequence multiple equivalent solutions are possible even when using UDRs. Considering that obtaining specific pairwise interaction data is often very resource-intensive, its inclusion, for this system, could not be justified.

Not only were there large variations in structures obtained using different sets of restraints, the results were also largely insensitive to the number of restraints. This suggests that performing multiple docking runs using different combinations of restraints and attempting to identify a consensus was more effective than applying all available restraints in a single run. Indeed, by default HADDOCK automatically removes a percentage of the restraints each trial. While this option is intended to account for possible errors in the data it was also effective when using ideal data as it allowed for the identification of a consensus solution.

Ultimately, the aim of many docking studies is to identify interactions that could be used to improve the potency and selectivity of a peptide for its drug target. For this, it is not so much the similarity between the proposed model and the actual co-crystal structure that is important but the ability to identify specific pairwise interactions that may be used to develop a potential mimetic or modify or enhance binding or selectivity. In this respect we note that in the system examined the maximum values for F_{nat}-PIR and F_{nat}-CIR for structures with an i-RMSD < 4 Å was about 80 % and 60 %, respectively, independent of the type or number of restraints used. Similarly, the maximum value for F_{nat} was about 50 % for these structures. A value of 50 % would be considered

to be a high accuracy prediction according to the CAPRI criteria but whether such values translate to a sufficiently predictive model will depend on the specific application and the question to be answered. For the system examined using the most favourable experimental scenario considered and multiple unambiguous as well as ambiguous restraints, the probability of finding a given structure within 4 Å was 60–70 %. The probability of correctly identifying a specific pairwise interaction in this structure was ~50 %. This meant that the overall probability of correctly identifying a specific pairwise interaction as a native contact in a docked structure was in the order of 30 %. This highlights the danger of considering just a single docked structure in isolation and the importance of examining the robustness of the solutions obtained. Finally we note that the results have been presented primarily in terms of averages over single structures (e.g. the top-ranked structure). All structures were however analysed. The trends in the probabilities obtained using all structures were essentially identical to those obtained using single structures. In a similar manner results associated with varying the number of restraints have been presented primarily for the C_AP_N trials, as experimentally this is the most realistic case. However, the trends observed using other combinations of input structures are very similar.

SUMMARY

Although widely used, the docking of flexible peptides and proteins remains highly challenging. Here the interaction between the venom peptide PcTx1 and the ion channel ASIC1 has been used as a case study to examine the effect of different combinations of restraints and input structures on the utility of the structures obtained. Using only information regarding which residues form the binding epitope of the peptide, the probability that the interface was well reproduced (i–RMSD < 4 Å) in the highest ranked or any other single structure was approximately 30 %. This probability increased to approximately 60 % when information on which channel residues are involved in binding was also included. The use of pairwise interaction data did not significantly improve the results. However, even using ideal data, a large degree of variability between runs was observed. The results were also very sensitive to the specific choice of input structures. For an experimentalist wishing to use a given docked structure to guide lead optimisation experiments, the probability of correctly identifying a specific pairwise interaction even in the most favourable scenario examined was approximately 30 %. Thus while the restraint-driven docking approach used in this work was capable of finding high quality solutions the average probability that a given solution (such as the top-ranked or most common structure) could be considered to be high quality in this case was low.

This underlines the importance of considering multiple solutions in docking studies and potentially using a statistical analysis of an ensemble of generated solutions to identify the interface. This would be especially important in cases where the alternative solutions cannot be distinguished based on the scoring function. It also highlights the practical distinction between developing docking methods and testing their accuracy and the use of structural models from docking approaches by experimentalists. In the first instance the interest is on whether one or more members of a set of possible set solutions is similar to a given experimental structure. For an experimentalist, however, the extent to which a given set of docked structure is accurate enough to plan future experiments is the primary interest. Finally, while this study was performed using a single peptide-channel complex the results should be representative of other systems where a small venom peptide is docked to a larger protein.

FIGURES AND TABLES

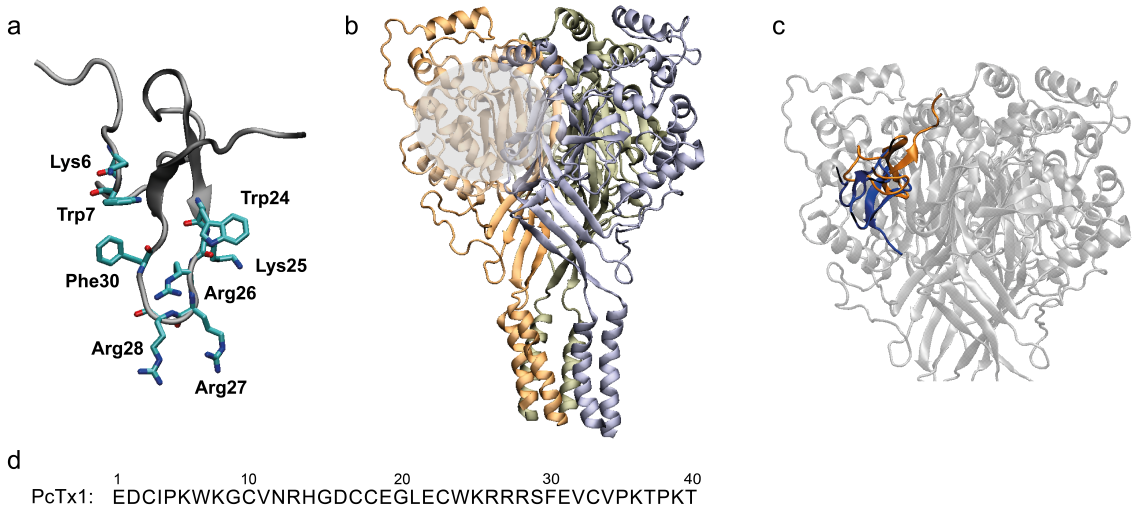

Figure 1. Structure of the peptide PcTx1, chicken ASIC1 (cASIC1), and the PcTx1-ASIC1 complex. **(a)** Cartoon representation of PcTx1 with the backbone and side chain of residues used for docking restraints shown explicitly. **(b)** Crystal structure of homotrimeric chicken ASIC1 (PDB 3S3X) [34] with the three subunits shown in orange, pale blue and pale green. The shaded region outlines the location of the acidic binding pocket formed at the interface of two channel subunits. **(c)** Extracellular domain of the PcTx1-cASIC1 crystal structure with an overlay of the peptide from docking simulations using experimental information on the peptide only.[25] The structure of the PcTx1-cASIC1 complex (PDB 3S3X) [34]with the channel shown in grey and PcTx1 shown in blue. Overlaid on this in orange is the orientation of PcTx1 obtained from docking simulations [25]. (d) The amino acid sequence of PcTx1.

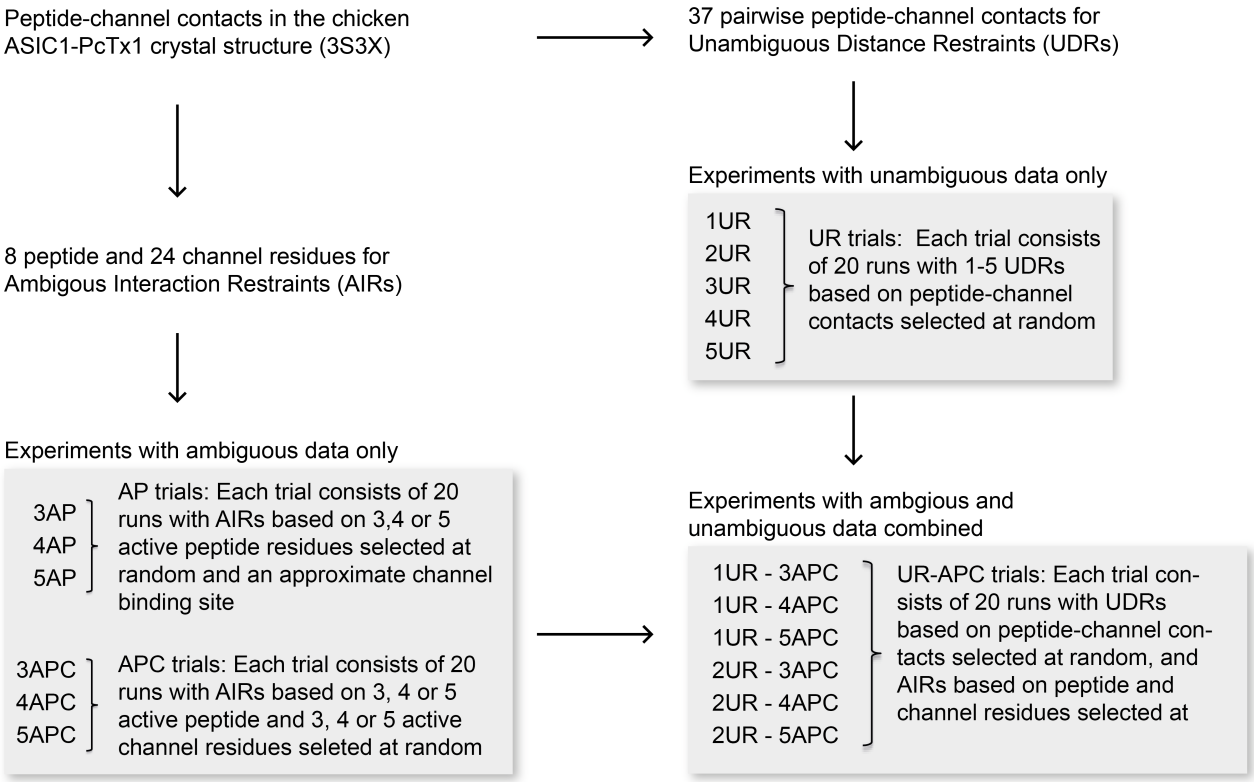

Figure 2. Overview of restraint-driven docking simulations. All restraints are based on the peptide-channel contacts at the complex interface in the PcTx1-cASIC1 crystal structure. [34] Four sets of docking simulations were carried out: (i) using ambiguous data based on binding information of peptide residues only (AP trials); (ii) using ambiguous data based on binding information of peptides and channel residues (APC trials); (iii) using unambiguous data only (UR trials); and (iv) using a combination of ambiguous and unambiguous data (UR-APC trials). The residues used for restraints are listed in Table S1 and S2 in the supporting information.

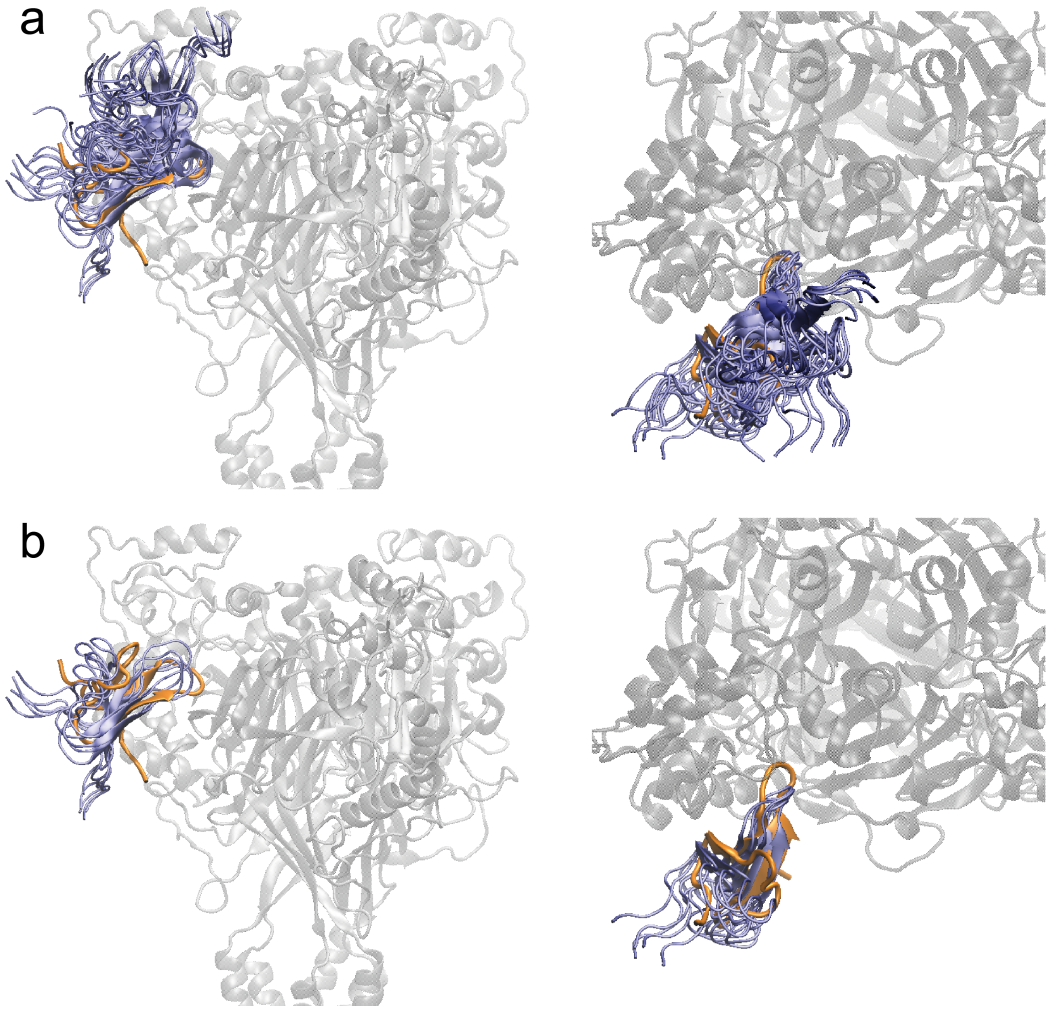

Figure 3. Overlay of top-ranked structures from docking runs in comparison to the co-crystal structure illustrating the large variations between runs with the same amount of restraints depending on the precise combination of residues used. Right-hand side images show a side-view (parallel to the membrane) of the entire extra-cellular domain of the channel. Left-hand side images show a top-view (perpendicular to the membrane) of the binding pocket on the channel. **(a)** Binding mode of PcTx1 predicted by the single top-ranked structures from the 20 runs of the C_AP_N-0UR-4AP trials (blue) in comparison to the peptide from the co-crystal structure (orange). **(b)** Binding mode of PcTx1 from 6 of the 20 runs in the C_AP_N-0UR-4AP trials in which the i-RMSD of the top-ranked structure is < 4.0 Å (blue) in comparison to the peptide from the co-crystal structure (orange).

Table 1. Results from docking simulations in which information on residues involved in binding is available for the peptide only (AP experiments). The table shows the probability (%) of the single top-ranking structure to be within a given interface RMSD (i-RMSD) cut-off of the co-crystal structure, averaged over the 20 independent runs in a trial. 3AP, 4AP and 5AP refer to the number of active peptide residues used to form ambiguous interaction restraints (AIRs). No unambiguous restraints (UR) were used. Different combinations of input structures for the peptide (P) and the channel (C) were used. C_X and C_A refer to the channel structure from the co-crystal structure and the apo-form of the channel. P_X and P_N refer to the peptide from the co-crystal structure or the consensus NMR structure. Note that 4Å i-RMSD would be considered acceptable according to the CAPRI standards. [12]

	1Å	**2 Å**	**3 Å**	**4 Å**	**5 Å**	**6 Å**
C_XP_X-0UR-3AP	35	50	50	50	65	65
C_XP_X-0UR-4AP	40	50	50	50	50	55
C_XP_X-0UR-5AP	40	45	45	45	65	70
C_XP_N-0UR-3AP	0	0	0	10	30	35
C_XP_N-0UR-4AP	0	0	5	15	25	30
C_XP_N-0UR-5AP	0	0	0	30	35	60
C_AP_N-0UR-3AP	0	0	10	20	35	45
C_AP_N-0UR-4AP	0	0	15	30	40	65
C_AP_N-0UR-5AP	0	0	10	20	35	55
C_AP_N-0UR-3AP$_{NR}$	0	0	10	20	25	50
C_AP_N-0UR-4AP$_{NR}$	0	0	15	20	25	60
C_AP_N-0UR-5AP$_{NR}$	0	0	45	45	55	70
		0-35 %		36-70 %		71-100 %

Table 2. Results from docking simulations in which information on residues involved in binding is available for the peptide and the channel (APC experiments). Results are given as the chances of finding a singe top-ranking structure within a given interface RMSD (i-RMSD) cut-off of the co-crystal structure, averaged over the 20 independent runs in a trial. 3AP, 4AP and 5AP refer to the number of active peptide and active channel residues used to form unambiguous interaction restraints (AIRs). No unambiguous restraints (UR) were used. Different combinations of input structures for the peptide (P) and the channel (C) were used. C_X and C_A refer to the channel structure from the co-crystal structure and the apo-form of the channel. P_X, P_N and P_{N20} refer to the peptide from the co-crystal structure, the consensus NMR structure or the ensemble of 20 NMR structures of PcTx1.

	i-RMSD					
	1Å	**2 Å**	**3 Å**	**4 Å**	**5 Å**	**6 Å**
C_XP_X-0UR-3APC	50	70	70	70	85	85
C_XP_X-0UR-4APC	40	60	60	60	75	75
C_XP_X-0UR-5APC	65	80	80	80	80	85
C_XP_N-0UR-3APC	0	0	10	10	20	55
C_XP_N-0UR-4APC	0	0	10	10	30	45
C_XP_N-0UR-5APC	0	0	0	20	40	55
C_XP_{N20}-0UR-3APC	0	0	15	50	50	65
C_XP_{N20}-0UR-4APC	0	0	40	40	50	65
C_XP_{N20}-0UR-5APC	0	0	30	40	40	70
C_AP_N-0UR-3APC	0	0	15	25	65	80
C_AP_N-0UR-4APC	0	0	45	60	75	90
C_AP_N-0UR-5APC	0	0	40	50	75	95
C_AP_N-0UR-3APC$_{NR}$	0	0	30	50	75	95
C_AP_N-0UR-4APC$_{NR}$	0	0	45	55	70	90
C_AP_N-0UR-5APC$_{NR}$	0	0	60	65	85	95
		0-35 %		36-70 %		71-100 %

Table 3: Results from docking simulations in which unambiguous data is used on its own (UR trials) or in combination with ambiguous data (UR-APC trials). Results are given as the chances of finding a single top-ranking structure within a given interface RMSD (i-RMSD) cut-off of the co-crystal structure, averaged over the 20 independent runs in a trial. 3AP, 4AP and 5AP refer to the number of active peptide and active channel residues used to form unambiguous interaction restraints (AIRs). 1, 2, 3,4 or 5 UR refers to the number of UDRs used. All runs used the apo-form of the channel (C_A) and the consensus NMR structure of PcTx1 (P_N).

	i-RMSD					
	1 Å	**2 Å**	**3 Å**	**4 Å**	**5 Å**	**6 Å**
C_AP_N-1UR	0	0	10	30	45	55
C_AP_N-2UR	0	0	20	45	65	75
C_AP_N-3UR	0	0	40	65	80	100
C_AP_N-4UR	0	0	40	60	75	95
C_AP_N-5UR	0	0	30	50	75	100
C_AP_N-1UR-3APC	0	0	20	55	60	85
C_AP_N-1UR-4APC	0	0	25	60	70	90
C_AP_N-1UR-5APC	0	0	45	80	85	100
C_AP_N-2UR-3APC	0	0	40	70	85	100
C_AP_N-2UR-4APC	0	0	40	65	70	95
C_AP_N-2UR-5APC	0	0	55	65	75	100
		0-35 %		36-70 %		71-100 %

Table 4: Results from four alternative approaches used to select a solution from the HADDOCK output. Data shown is for APC experiments.

	i-RMSD					
	1 Å	**2 Å**	**3 Å**	**4 Å**	**5 Å**	**6 Å**
% of runs where the top-ranked structure has an i-RMSD < cut-off						
Top-scored structure						
C_AP_N-0UR-3APC	0	0	15	25	65	80
C_AP_N-0UR-4APC	0	0	45	60	75	90
C_AP_N-0UR-5APC	0	0	40	50	75	95
% of runs where the most common structure has an i-RMSD < cut-off						
C_AP_N-0UR-3APC	0	0	10	35	60	85
C_AP_N-0UR-4APC	0	0	30	50	55	85
C_AP_N-0UR-5APC	0	0	65	85	90	100
% of runs with at least one structure from the 200 water-refined structures with i-RMSD < cut-off						
C_AP_N-0UR-3APC	0	0	7	26	36	53
C_AP_N-0UR-4APC	0	0	11	32	41	59
C_AP_N-0UR-5APC	0	0	20	47	58	74
% of runs with at least one structure from the best 4 structures in top-ranked cluster with i-RMSD < cut-off						
C_AP_N-0UR-3APC	0	0	15	30	59	85
C_AP_N-0UR-4APC	0	0	40	54	71	90
C_AP_N-0UR-5APC	0	0	38	59	85	100
		0-35 %		36-70 %		71-100 %

SUPPORTING INFORMATION

The following additional Tables and Figures can be found in Supporting Information. Figure S1 shows an overlay of docked structure to the co-crystal structure with i-RMSD values from 1 to 6 Å. Table S1 lists the peptide and channel residues used as input for the ambiguous interaction restraints. Table S2 lists the peptide-channel residues pairs used as input for unambiguous distance restraints. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*co-corresponding authors.

Prof. Alan E. Mark: Phone: +61 7 336 54180; Email: a.e.mark@uq.edu.au, or

Dr. Evelyne Deplazes, Phone: +61 7 336 57562 ; Email: e.deplazes@uq.edu.au

Author Contributions

E.D, A.E.M and G.K conceived the idea and designed the experiments. E.D and J.D. performed the docking runs. E.D, A.E.M and A.M.J.J.B. analyzed and interpreted the data. All authors contributed to the writing of the manuscript and have approved the final version.

Funding Sources

E.D was supported by a Fellowship from the Swiss National Science Foundation and currently holds an Early-Career Research Fellowship from the Australian National Health & Medical Research Council (NHMRC). A.E.M. is an ARC Discovery Outstanding Researcher. G.F.K. is a NHMRC Principal Research Fellowship. J.D. was supported by an Advantage summer research scholarship by the University of Queensland. The docking runs were carried out on the HADDOCK web-server, which is supported by the WeNMR project. The WeNMR project (European FP7 e-

Infrastructure grant, contract no. 261572, www.wenmr.eu), supported by the European Grid Initiative (EGI) through the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, South Africa, Malaysia, Taiwan, the Latin America GRID infrastructure via the Gisela project and the US Open Science Grid (OSG) are acknowledged for the use of web portals, computing and storage facilities.

NOTES

The authors declare no competing financial interest.

REFERENCES

1. King, G. F., Venoms as a Platform for Human Drugs: Translating Toxins into Therapeutics. *Expert Opin. Biol. Ther.* **2011**, 11, 1469-1484.
2. Saez, N. J.; Senff, S.; Jensen, J. E.; Er, S. Y.; Herzig, V.; Rash, L. D.; King, G. F., Spider-Venom Peptides as Therapeutics. *Toxins* **2010**, 2, 2851-2871.
3. Chen, P. C.; Kuyucak, S., Developing a Comparative Docking Protocol for the Prediction of Peptide Selectivity Profiles: Investigation of Potassium Channel Toxins. *Toxins* **2012**, 4, 110-38.
4. Chen, R.; Robinson, A.; Gordon, D.; Chung, S.-H., Modeling the Binding of Three Toxins to the Voltage-Gated Potassium Channel (Kv1.3). *Biophys. J.* **2011**, 101, 2652-2660.
5. Dutertre, S.; Nicke, A.; Lewis, R. J., B2 Subunit Contribution to 4/7 A-Conotoxin Binding to the Nicotinic Acetylcholine Receptor. *J. Biol. Chem.* **2005**, 280, 30460-30468.
6. Dutertre, S.; Nicke, A.; Tyndall, J. D. A.; Lewis, R. J., Determination of A-Conotoxin Binding Modes on Neuronal Nicotinic Acetylcholine Receptors. *J. Mol. Recognit.* **2004**, 17, 339-347.
7. Eriksson, M. A.; Roux, B., 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.* **2002**, 83, 2595-2609.
8. Jin, L.; Wu, Y., Molecular Mechanism of the Sea Anemone Toxin ShK Recognizing the Kv1.3 Channel Explored by Docking and Molecular Dynamic Simulations. *J. Chem. Inf. Model.* **2007**, 47, 1967-1972.
9. Pennington, M. W.; Harunur Rashid, M.; Tajhya, R. B.; Beeton, C.; Kuyucak, S.; Norton, R. S., A C-terminally Amidated Analogue of ShK Is a Potent and Selective Blocker of the Voltage-Gated Potassium Channel Kv1.3. *FEBS Letters* **2012**, 586, 3996-4001.
10. Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J., Principles of Flexible Protein–Protein Docking. *Proteins: Struct. Funct. Bioinform.* **2008**, 73, 271-289.
11. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**, 3, 935-949.
12. Lensink, M. F.; Méndez, R.; Wodak, S. J., Docking and Scoring Protein Complexes: CAPRI 3rd Edition. *Proteins: Struct. Funct. Bioinform.* **2007**, 69, 704-718.
13. Lensink, M. F.; Wodak, S. J., Docking, Scoring, and Affinity Prediction in CAPRI. *Proteins: Struct. Funct. Bioinform.* **2013**, 81, 2082-2095.
14. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J., Protein–Protein Docking Dealing with the Unknown. *J. Comput. Chem.* **2010**, 31, 317-342.
15. Yuriev, E.; Agostino, M.; Ramsland, P. A., Challenges and Advances in Computational Docking: 2009 in Review. *J. Mol. Recognit.* **2011**, 24, 149-164.
16. Yuriev, E.; Ramsland, P. A., Latest Developments in Molecular Docking: 2010–2011 in Review. *J. Mol. Recognit.* **2013**, 26, 215-239.
17. Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, 125, 1731-1737.
18. Karaca, E.; Melquiond, A. S. J.; de Vries, S. J.; Kastritis, P. L.; Bonvin, A. M. J. J., Building Macromolecular Assemblies by Information-Driven Docking: Introducing the HADDOCK Multibody Docking Server. *Mol. Cell. Proteomics* **2010**, 9, 1784-1794.
19. Schmitz, C.; Melquiond, A. S. J.; de Vries, S. J.; Karaca, E.; van Dijk, M.; Kastritis, P. L.; Bonvin, A. M. J. J. Protein–Protein Docking with HADDOCK. In *NMR of*

Biomolecules: Towards Mechanistic Systems Biology; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2012, pp 520-535.
20. van Ingen, H.; Bonvin, A. M. J. J., Information-Driven Modeling of Large Macromolecular Assemblies Using NMR Data. *J. Magn. Reson.* **2014**, 241, 103-114.
21. Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J., CAPRI: A Critical Assessment of Predicted Interactions. *Proteins* **2003**, 52, 2-9.
22. Rodrigues, J. P. G. L. M.; Bonvin, A. M. J. J., Integrative Computational Modeling of Protein Interactions. *FEBS J.* **2014**, 281, 1988-2003.
23. Lefèvre, F.; Rémy, M.-H.; Masson, J.-M., Alanine-Stretch Scanning Mutagenesis: A Simple and Efficient Method to Probe Protein Structure and Function. *Nucleic Acids Res.* **1997**, 25, 447-448.
24. Wells, J. A. Systematic Mutational Analyses of Protein-Protein Interfaces. In *Methods Enzymol.*, John, J. L., Ed.; Academic Press: New York, 1991; Vol. Volume 202, pp 390-411.
25. Saez, N. J.; Mobli, M.; Bieri, M.; Chassagnon, I. R.; Malde, A. K.; Gamsjaeger, R.; Mark, A. E.; Gooley, P. R.; Rash, L. D.; King, G. F., A Dynamic Pharmacophore Drives the Interaction between Psalmotoxin-1 and the Putative Drug Target Acid-Sensing Ion Channel 1a. *Mol. Pharmacol.* **2011**, 80, 796-808.
26. Chu, X.; Papasian, C.; Wang, J.; Xiong, Z., Modulation of Acid-Sensing Ion Channels: Molecular Mechanisms and Therapeutic. *Int. J. Physiol. Pathophysiol. Pharmacol.* **2011**, 3, 288-309.
27. Leng, T. D.; Xiong, Z. G., The Pharmacology and Therapeutic Potential of Small Molecule Inhibitors of Acid-Sensing Ion Channels in Stroke Intervention. *Acta Pharmacol. Sin.* **2013**, 34, 33-38.
28. Sluka, K. A.; Winter, O. C.; Wemmie, J. A., Acid-Sensing Ion Channels: A New Target for Pain and CNS Diseases. *Curr. Opin. Drug Discov. Devel.* **2009**, 12, 693-704.
29. Wemmie, J. A.; Taugher, R. J.; Kreple, C. J., Acid-Sensing Ion Channels in Pain and Disease. *Nat. Rev. Neurosci.* **2013**, 14, 461-471.
30. Lingueglia, E.; Lazdunski, M., Pharmacology of ASIC Channels. *Wiley Interdiscip. Rev. Membr. Transp. Signal.* **2013**, 2, 155-171.
31. Escoubas, P.; De Weille, J. R.; Lecoq, A.; Diochot, S.; Waldmann, R.; Champigny, G.; Moinier, D.; Ménez, A.; Lazdunski, M., Isolation of a Tarantula Toxin Specific for a Class of Proton-Gated Na^+ Channels. *J. Biol. Chem.* **2000**, 275, 25116-25121.
32. Salinas, M.; Rash, L. D.; Baron, A.; Lambeau, G.; Escoubas, P.; Lazdunski, M., The Receptor Site of the Spider Toxin PcTx1 on the Proton-Gated Cation Channel ASIC1a. *J. Physiol.* **2006**, 570, 339-354.
33. Baconguis, I.; Gouaux, E., Structural Plasticity and Dynamic Selectivity of Acid-Sensing Ion Channel-Spider Toxin Complexes. *Nature* **2012**, 489, 400-405.
34. Dawson, R. J. P.; Benz, J.; Stohler, P.; Tetaz, T.; Joseph, C.; Huber, S.; Schmid, G.; Hügin, D.; Pflimlin, P.; Trube, G.; Rudolph, M. G.; Hennig, M.; Ruf, A., Structure of the Acid-Sensing Ion Channel 1 in Complex with the Gating Modifier Psalmotoxin 1. *Nat. Commun.* **2012**, 3, 1-8.
35. Pietra, F., Docking and MD Simulations of the Interaction of the Tarantula Peptide Psalmotoxin-1 with ASIC1a Channels Using a Homology Model. *J. Chem. Inf. Model.* **2009**, 49, 972-977.
36. Qadri, Y. J.; Berdiev, B. K.; Song, Y.; Lippton, H. L.; Fuller, C. M.; Benos, D. J., Psalmotoxin-1 Docking to Human Acid-Sensing Ion Channel-1. *J. Biol. Chem.* **2009**, 284, 17625-17633.
37. de Vries, S. J.; van Dijk, M.; Bonvin, A. M. J. J., The HADDOCK Web Server for Data-Driven Biomolecular Docking. *Nat. Protoc.* **2010**, 5, 883-897.

38. van Dijk, A. D.; Bonvin, A. M., Solvated Docking: Introducing Water into the Modelling of Biomolecular Complexes. *Bioinformatics (Oxford, England)* **2006**, 22, 2340-2347.
39. Jasti, J.; Furukawa, H.; Gonzales, E. B.; Gouaux, E., Structure of Acid-Sensing Ion Channel 1 at 1.9 Å Resolution and Low pH. *Nature* **2007**, 449, 316-323.
40. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999**, 38, 236-240.
41. Hwang, H.; Vreven, T.; Janin, J.; Weng, Z., Protein–Protein Docking Benchmark Version 4.0. *Proteins: Structure, Function, and Bioinformatics* **2010**, 78, 3111-3114.

Table of Content Graphics