

COMPUTATIONAL MODELING OF PHOSPHOTRANSFER COMPLEXES IN TWO-COMPONENT SIGNALING

Alexander Schug,^{*} Martin Weigt,[†] James A. Hoch,[‡]
Jose N. Onuchic,^{*} Terence Hwa,^{*} and Hendrik Szurmant[‡]

Contents

1. Introduction	44
2. Methods	47
2.1. Structure-based simulations	47
2.2. Relaxation in an empirical force field	54
3. Summary	54
Acknowledgments	55
References	55

Abstract

Two-component signal transduction systems enable cells in bacteria, fungi, and plants to react to extracellular stimuli. A sensor histidine kinase (SK) detects such stimuli with its sensor domains and transduces the input signals to a response regulator (RR) by *trans*-phosphorylation. This *trans*-phosphorylation reaction requires the formation of a complex formed by the two interacting proteins. The complex is stabilized by transient interactions. The nature of the transient interactions makes it challenging for experimental techniques to gain structural information. X-ray crystallography requires stable crystals, which are difficult to grow and stabilize. Similarly, the mere size of these systems proves problematic for NMR. Theoretical methods can, however, complement existing data. The statistical direct coupling analysis presented in the previous chapter reveals the interacting residues at the contact interface of the SK/RR pair. This information can be combined with the structures of the individual proteins in molecular dynamical simulation to generate structural models of the complex. The general approach, referred to as MAGMA, was tested on the

^{*} Center for Theoretical Biological Physics, University of California San Diego, La Jolla, California, USA

[†] Institute for Scientific Interchange, Viale S. Severo 65, Torino, Italy

[‡] Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California, USA

sporulation phosphorelay phosphotransfer complex, the Spo0B/Spo0F pair, delivering crystal resolution accuracy. The MAGMA method is described here in a step-by-step explanation. The developed parameters are transferrable to other SK/RR systems.

1. INTRODUCTION

Two-component signal transduction systems (TCS) enable bacteria, fungi, and plants to respond to stimuli and changes of environments like nutrients, light, or pressure (for a recent review see [Mascher *et al.*, 2006](#)). They consist of two proteins, a sensor histidine kinase (SK) and a response regulator (RR). The multidomain SK consists of intra- and/or extracellular sensor domains and a catalytic histidine kinase core (reviewed in [Szurmant *et al.*, 2007](#)). The latter can be subdivided into the phosphorylatable histidine containing HisKA domain and the catalytic ATP-binding ATPase domain. In response to a stimulus, the phosphoryl flux between the SK and the RR is modulated. As a first step, the histidine on the HisKA domain is autophosphorylated. In a second step, this phosphoryl group is transferred to an aspartate residue on the RR protein, which most commonly serves as a transcription factor ([Galperin, 2006](#)).

Extended versions of the TCS signal transduction pathway are the phosphorelays ([Hoch, 2000](#)). In these systems, the phosphoryl flux between the SK and the RR is mediated by a second single domain RR and a phosphotransferase in a His-Asp-His-Asp phosphotransfer cascade. The phosphotransferase protein can feature either a monomeric four-helix bundle Hpt domain (also utilized by the chemotaxis histidine kinase, CheA) or a second dimeric four-helix bundle HisKA-like domain ([Hoch and Varughese, 2001](#)). The latter form of phosphorelay is exemplified by the well-described sporulation phosphorelay of the *Bacilli*, which connects five SK KinA–KinE with the sporulation RR Spo0A via the single domain RR Spo0F and the HisKA-like phosphotransferase Spo0B ([Burbulys *et al.*, 1991](#); [Jiang *et al.*, 2000](#)).

The phosphotransfer complex between HisKA and RR domains is ruled by transient interactions. While many individual TCS proteins have been structurally resolved, these transient interactions result in short-lived and unstable complexes, which have proven resistant to structural resolution by experimental means. For this reason, as of August, 2009, the complex of above described Spo0F and Spo0B proteins of the *Bacillus subtilis* sporulation phosphorelay remains as the only published structural example of a HisKA–RR pair, trapped in the act of phosphotransfer ([Zapf *et al.*, 2000](#)).

This lack of structural templates hampers theoretical structure–prediction methods. Established methods like homology modeling ([Eswar *et al.*, 2008](#))

rely on the presence of such templates with high sequential similarity. Physics-based approaches (Fujitsuka *et al.*, 2004; Hardin *et al.*, 2002; Schug *et al.*, 2005a) struggle with the accuracy of the underlying force-fields, especially for mixed α -helical/ β -sheet structures (Best *et al.*, 2008). Even assuming perfect suitability of a specific force-field, the mere size of TCS protein complexes might make such prediction methods computationally prohibitive by requiring searches of huge conformational spaces with accordingly high numbers of degrees of freedom (“problem of sampling”) (Schug and Wenzel, 2006; Schug *et al.*, 2005b). Currently, no reliable “gold standard” for the prediction of protein complexes has been found as demonstrated by the CAPRI competitions, in which different approaches at protein–protein docking are compared (Janin *et al.*, 2003).

Not all is bad news, however. The current ongoing “genomic revolution” provides scientists with a wealth of sequential information and complete genomes for an exponentially growing number of systems. As TCS are ubiquitously used and highly amplified in bacteria, fungi, and plants, one can take advantage of the abundance of sequential information. The *direct coupling analysis* (DCA) (Weigt *et al.*, 2009), presented in the previous chapter, investigates the mutational patterns of coevolving protein like the SK/RR pair in TCS and has three subsequent steps. The first step is a homology search of the two target sequences in as many bacterial genomes as possible. The second step constitutes aligning the sequences and performing a covariance analysis. This reveals pairs of amino acids with high mutual information in the two proteins. In a third step, a statistical inference analysis distinguishes between pairs, which directly interact, from such, which are indirectly correlated, that is, for which the correlation is mediated by additional residue positions. These three steps result in a set of directly interacting pairs of amino acids between the SK and the RR, that is residue positions, which lie on the interface of the two proteins (here thereto referred as DI contacts). This information describes the evolutionary most crucial interactions of the surface and is therefore by itself of high scientific interest. It becomes, however, even more valuable when combined with molecular dynamical (MD) simulations.

MD-simulations approximate the physics of an entire system of interacting biomolecules over a defined period of time. Typically, one has to balance the accuracy of description/physics with the available computational resources (Adcock and McCammon, 2006). Here, we employ native structure-based simulations (SBS), which are highly successful in describing protein folding (Clementi *et al.*, 2001; Onuchic and Wolynes, 2004; Onuchic *et al.*, 1997), conformational transitions related to protein function (Schug *et al.*, 2007; Whitford *et al.*, 2007), RNA folding and function (Thirumalai and Hyeon, 2005; Whitford *et al.*, 2009b), and protein–protein interactions (Levy and Onuchic, 2006; Levy *et al.*, 2007), and make

prediction in agreement with experimental measurements (Clementi and Plotkin, 2004; Gambin *et al.*, 2009). They describe protein/RNA dynamics based on a specific structural conformation, which is usually the native state. This results in a concise Hamiltonian and allows adopting the level of coarse-graining to the individual scientific question (Lammert *et al.*, 2009; Oliveira *et al.*, 2008; Schug *et al.*, 2009a; Whitford *et al.*, 2009a).

We previously demonstrated that combining complementary independent information, the coevolutionary information obtained from DCA and structural data of the unbound monomers, by SBS can predict a TCS complex in agreement with experimental data (see Fig. 3.1) (Schug *et al.*, 2009b). The approach will hereafter be referred to as MAGMA (Molecular dynamics And Genomic information for Macromolecular Assembly). The MAGMA method along with some of its results is described here in some detail. Relying on the DCA analysis of subtle *mutational patterns* is an orthogonal approach to typical structure-prediction methods like homology modeling, which rely on *highly conserved residues* in sequential and structural libraries. The ultimate aim is, however, not stopping at protein structure prediction. Right now, we only use DCA to determine interacting residues

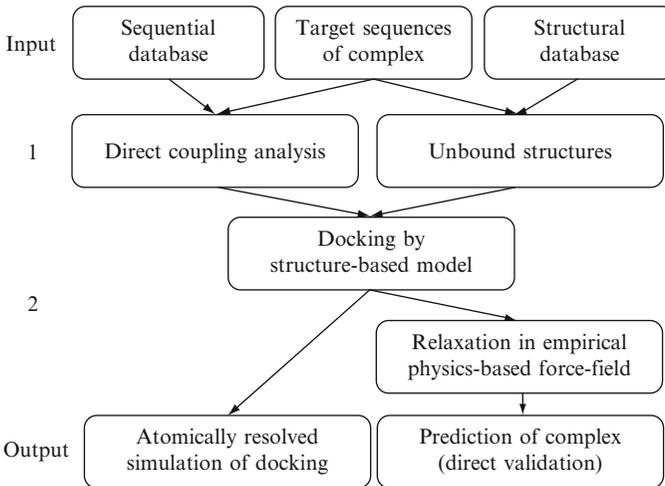


Figure 3.1 Flow-chart of the MAGMA approach. Given the target sequence of an unknown protein complex, direct-coupling analysis (DCA, see previous chapter) investigates statistical fluctuations of mutational pattern in sequential homologues and suggests pairwise contacts defining an interaction surface. Similarly, unbound structures for the given target sequences can be either directly extracted from a structural database or generated by structure-prediction methods like homology modeling. This information of the unbound structures and interaction surface contacts is sufficient information for docking simulations in computationally efficient structure-based models, providing both insight into the mechanism of docking and making a prediction of the protein complex. To improve the quality of the prediction, it can be additionally relaxed in physics-based empirical force fields.

at the intermolecular docking interface of RR and SK. We hope to combine the additional DCA intramolecular information with molecular simulations to generate realistic descriptions of the conformational changes underlying the biological function of biomolecules. In the context of TCS, that would be the simulation of conformational changes during autophosphorylation, docking of the RR to the SK, phosphoryl-transfer, and the final dissociation of RR and SK in the absence of detailed experimental structural data for each step.

2. METHODS

2.1. Structure-based simulations

Structure-based simulations are used in protein folding simulations based on the funneled energy landscape and the principle of minimal frustration. Accordingly, evolution shapes and concurrently smoothens the energy landscape of proteins by ensuring a dominance of interactions present in the native state during the entire folding process (Bryngelson *et al.*, 1995; Frauenfelder *et al.*, 1991; Onuchic and Wolynes, 2004). This guiding bias prevents entrapment in minima representing nonnative folds. It also provides a degree of robustness, permitting protein folding and function despite moderate environmental changes or mutations.

In folding simulations, native structure-based models¹ represent the ideal case of a perfectly funneled energy landscape where only interactions present in the native state are taken into account and no energetic frustration occurs. They have shown to be in high agreement with experimental measurements (Chavez *et al.*, 2004; Cheung *et al.*, 2003; Clementi *et al.*, 2000). In a typical mathematical description, each amino acid is represented as a single C_α -bead. Bridging these methods toward empirical all-atom force fields, variants using multiwelled Gaussians for the contacts (Lammert *et al.*, 2009), $C_\alpha C_\beta$ (Finke *et al.*, 2004; Oliveira *et al.*, 2008) or all-atom representations have been developed (Linhananta and Zhou, 2002; Shimada *et al.*, 2001; Whitford *et al.*, 2009a; Zhou *et al.*, 2003). As the latter incorporate the details of packing best while maintaining computational tractability, we choose (Whitford *et al.*, 2009a) as a basis for our docking simulations (the Hamiltonian is given as Eq. (3.1)):

¹ Structure-based models are often referred to as Go-models.

$$\begin{aligned}
E = & \sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{impropers/planar}} K_\chi (\chi - \chi_0)^2 \\
& + \sum_{\text{backbone}} K_{\text{BB}} F_{\text{D}}(\varphi)^2 + \sum_{\text{sidechain}} K_{\text{SC}} F_{\text{D}}(\varphi)^2 \\
& + \sum_{\text{contacts}} \left\{ \varepsilon_{\text{C}} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \varepsilon_{\text{NC}} \left(\frac{\sigma_{\text{NC}}}{r_{ij}} \right)^{12} \right\}, \\
F_{\text{D}}(\varphi) = & [1 - \cos(\varphi - \varphi_0)] + \frac{1}{2} [1 - \cos(3(\varphi - \varphi_0))] \tag{3.1}
\end{aligned}$$

with $K_r = 100k_{\text{B}}T (\text{\AA}^2)$, $K_\theta = 20k_{\text{B}}T$, $K_\chi = 20k_{\text{B}}T$, and $\varepsilon_{\text{NC}} = 0.01k_{\text{B}}T$. The values for r_0 , θ_0 , χ_0 , Φ_0 , and σ_{ij} are given by the native conformation. An illustration of the different terms can be found in [Figs. 3.2 and 3.3](#). Both impropers/planar and backbone denote dihedral terms, pending on how rigid they are in the structure (for a thorough discussion see [Oliveira *et al.*, 2008](#)). i and j run over all atoms and r_{ij} is the distance between any two atoms. An attractive interaction with $\varepsilon_{\text{C}}(i, j) = 1k_{\text{B}}T$ and $\varepsilon_{\text{NC}}(i, j) = 0k_{\text{B}}T$ is assigned to natively interacting residues, while $\varepsilon_{\text{C}}(i, j) = 1k_{\text{B}}T$ and $\varepsilon_{\text{NC}}(i, j) = 0.01k_{\text{B}}T$ enforce an excluded volume for noninteracting residues. σ_{ij} is the native distance between interacting residues and set to 2.5 Å for noninteracting residues. The dihedral strengths K_{BB} and K_{SC} are assigned in a way that the interaction energy between the sidechains and the backbones is balanced 2:1 and the total contacts energy (determined by the total number of contacts) is balanced 2:1 against the total sidechain energy ([Whitford *et al.*, 2009a](#)).

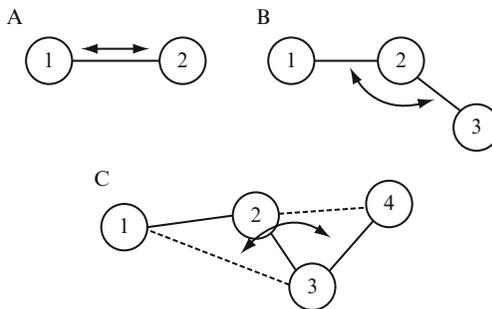


Figure 3.2 Illustration of the different interactions. The backbone interactions of native structure-based simulations have three contributions. A harmonic pair interaction (A) involves two atoms and describes the vibrations around a harmonic bond. The angle term or 1–3 interaction (B) is given by the angle between the bonds 1–2 and 2–3. Finally, the dihedral term (C) or 1–4 interaction is defined as the angle between two planes. In the example, atoms 1–3 span the first plane and atoms 2–4 the second plane.

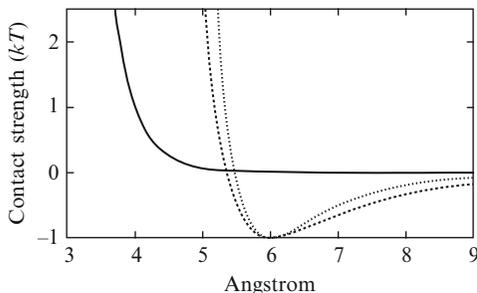


Figure 3.3 Illustration of the contact potential. The Lennard-Jones-type contact potential of structure-based models biases against distances much shorter than the contact distance σ_{ij} while providing an attractive basin for values around the distance. This results from the repulsive $(\sigma_{ij}/r_{ij})^{12}$ -term, which dominates for short distances (solid line, $\sigma_{ij} = 4 \text{ \AA}$) and prevents overlap of the electron shells of any two atoms. Typically, structure-based models have an attractive $(\sigma_{ij}/r_{ij})^6$ or $(\sigma_{ij}/r_{ij})^{10}$ -term for natively interacting residues (dashed lines) with the former being “softer” (longer bars) and the latter being more localized (shorter bars).

2.1.1. Deriving parameters for molecular docking simulations

SBS allow direct modeling of the independent protein monomers, which shall be docked. For docking, however, we need to introduce *a priori* unknown interprotein forces. A weak harmonical center-of-mass force for all atoms ($k = 0.25 \times 10^{-6} k_B T, \text{ \AA}^2$) simulates a sufficient molecular concentration of both molecules to bring the proteins into frequent contact with each other (Schug *et al.*, 2007).

Another crucial term in MAGMA is the specific inclusion of the DCA-predicted residue pairs (see previous chapter and Weigt *et al.*, 2009). These predicted direct interactions at the surface will determine the exact orientation of the two molecules with respect to each other (“docking pose”). The most natural implementation of these interactions is as additional contacts. To prevent any bias or artifacts from overfitting to a dataset, we choose a homogenous distance for these contacts between the corresponding C_α atoms. As typically two amino acids, which are in contact with each other interact by 3–6 contacts between individual atoms on the all-atom level, the contact strength between the C_α atoms should account for the total interaction between the amino acids and was hence increased fivefold. We tested contact distances between 5.5 and 7 \AA , which is the range of average contact distances² over a range of reduced temperatures (1/3–2/3), all well below folding temperature to ensure fast convergence to a docked complex (“kinetic simulations”).

² It is important not to overestimate the contact distance. Otherwise the repulsive part of the van der Waals contact potential will dominate the interaction and prevent close approach of the interacting residues (see Fig. 3.3).

2.1.2. Test: Docking of the Spo0B/Spo0F complex

To develop and test our parameters, we choose the described Spo0B/Spo0F system, since both, structures of the individual proteins (pdb-codes: 1pey, Mukhopadhyay *et al.*, 2004; and 1ixm, Varughese *et al.*, 1998) and a structure of the complexed crystal (1f51; Zapf *et al.*, 2000) have been determined (see also Figs. 3.4 and 3.5).

In the case of several copies of a protein in a pdb-file, for example, as a result from crystal packing, one has to choose a representative conformation. We choose as a consistent but somewhat arbitrary choice the first representation in each pdb-file. The two proteins are then combined into one pdb-file with consecutively numbered amino acids and atoms.³ It is important to check that the two proteins do not overlap, as the resulting atomic clashes will stop the MD-simulations. If the proteins overlap or if one wants to speed up the subsequent docking simulations, the two proteins can be brought into spatial vicinity of each other with the docking interfaces close to each other and without overlap of the atoms by using, for example, the VMD software (Humphrey *et al.*, 1996) (Mouse → Move → Fragment and Mouse → Rotate).



Figure 3.4 The Spo0B/Spo0F complex. The phosphotransferase Spo0B/Spo0F system is part of a phosphorelay in the sporulation pathway of *Bacillus subtilis*. (Left) The crystal structure (PDB-ID 1f51) shows Spo0B (light gray) and Spo0F (dark gray). The residues His30 and Asp54 responsible for the phosphoryl-transfer are highlighted in black. (Right) DCA identifies 6 residue pairs, which are highly directly correlated (black) at the interface of the two proteins. Docking simulations using this information, information about the spatial vicinity of the His-Asp pair (also black), and the unbound protein structures generate a structural model in high agreement with the experimental structure ($C\alpha$ -RMSD < 3 Å).

³ The consecutive numbering avoids ambiguities, for example, when software packages ignore the chain identifier.

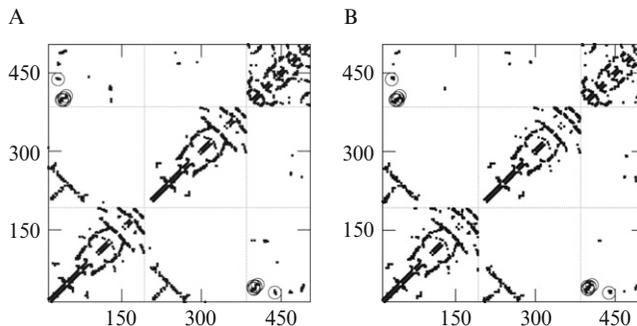


Figure 3.5 Contact maps of the Spo0B/Spo0F complex. The contact maps of (left) the crystal structure and (right) the prediction agree well. The axes denote the consecutively numbered residues of Spo0B-A (1–192), Spo0B-B (193–384), and Spo0F (385+). The DCA-contacts and the Asp-His pair are highlighted as circles. The quadrant, which contains the contacts of the interface region of the complex shows in addition to these explicitly included contacts other contacts formed in the crystal structure which have not been found by DCA. Some of these “missing” contacts are reconstituted in our docking simulations. It seems therefore possible that this subset of contacts consists of the crucial contacts for docking based on two facts: (A) DCA identifies them to have statistically strongly linked coevolution compared to all possible interface contact pairs and (B) they are sufficient information for successful reconstitution of the protein complex in docking by MAGMA.

The next step is preparing the docking simulations. Here, we use the GROMACS software package (Kutzner *et al.*, 2007; Van Der Spoel *et al.*, 2005). The required files for the simulations can be created by, for example, the webpage <http://sbm.ucsd.edu> (prepare a simulation, default parameters except contact map: Cut-off instead of shadow-map; the same webpage also contains a small tutorial into structure-based simulations). One receives two files, the gro-file with the atomic coordinates and a top-file with simulation parameters. In order to add the DCA-contacts, one has to edit the top-file using a text-editor right after the [pairs]-entry. The format is:

```
[pairs]
i j  $C(\sigma_{ij})^6$   $C(\sigma_{ij})^{12}$ 
```

i and *j* are the atom numbers between which a Lennard-Jones-type contact potential is introduced, σ_{ij} designates the desired contact distance in nm (we suggest 0.7 nm, see below), and *C* is the contact strength. For Spo0B/Spo0F, DCA predicts six contacts with significant direct information (see Table 3.1) (Weigt *et al.*, 2009).⁴ The crucial His30–Asp54 interaction

⁴ As discussed in the previous chapter the top 10 DI pairings are contacts that could have been included for docking analysis. For the present analysis only the six high DI contacts that also showed above threshold mutual information, as published in Weigt *et al.* (2009) were considered. We do not anticipate that results would change much if the additional four pairings are included since they involve the same response regulator and sensor kinase residues.

Table 3.1 Variation of C_α - C_α contact distances in docking simulations of Spo0F/Spo0B

Spo0B	Spo0F	Native distance (Å)	C_α - C_α distance (Å), 6 Å	C_α - C_α distance (Å), 6.5 Å	C_α - C_α distance (Å), 7 Å	C_α - C_α distance (Å), 7/11 Å
GLN 37	ILE 15	7.7	7.4 ± 0.2	7.4 ± 0.1	7.8 ± 0.2	7.7 ± 0.1
LEU 38	GLY 14	6.7	6.0 ± 0.2	6.4 ± 0.1	6.8 ± 0.1	6.9 ± 0.1
GLY 41	LEU 18	7.0	6.0 ± 0.1	6.5 ± 0.1	6.9 ± 0.2	6.9 ± 0.1
ASN 42	GLY 14	9.7	6.0 ± 0.1	6.5 ± 0.1	7.0 ± 0.1	6.9 ± 0.1
ASN 42	LEU 18	7.0	5.9 ± 0.1	6.4 ± 0.1	7.0 ± 0.1	6.9 ± 0.1
LEU 45	VAL 22	8.2	6.0 ± 0.2	6.5 ± 0.1	6.9 ± 0.1	7.8 ± 0.2
HIS 30	ASP 54	12.2	16.5 ± 0.5	15.9 ± 0.4	16.4 ± 0.3	11.6 ± 0.1

When varying the C_α - C_α distance of the DI contacts (first six rows) and the His-Asp pair (last row) in docking simulations, the deviations to the native distances are small for all choices of parameters. Each number represents 10 docking simulations at $T = 1/3$ with contact strengths for the DI predicted contacts of $5kT$. The parameters in the last column have an additional contact added between His30 and Asp54 of $10kT$.

responsible for the phosphoryl transfer cannot be detected by DCA due to perfectly conserved amino acids (covariance needs variance). We therefore test including an additional contact (see Fig. 3.6 and Table 3.1). To accommodate for the size of the phosphoryl, we add 4 Å to the contact distance, close to the typical size of a phosphate group in empirical force fields. This additional contact improves the quality of the prediction.

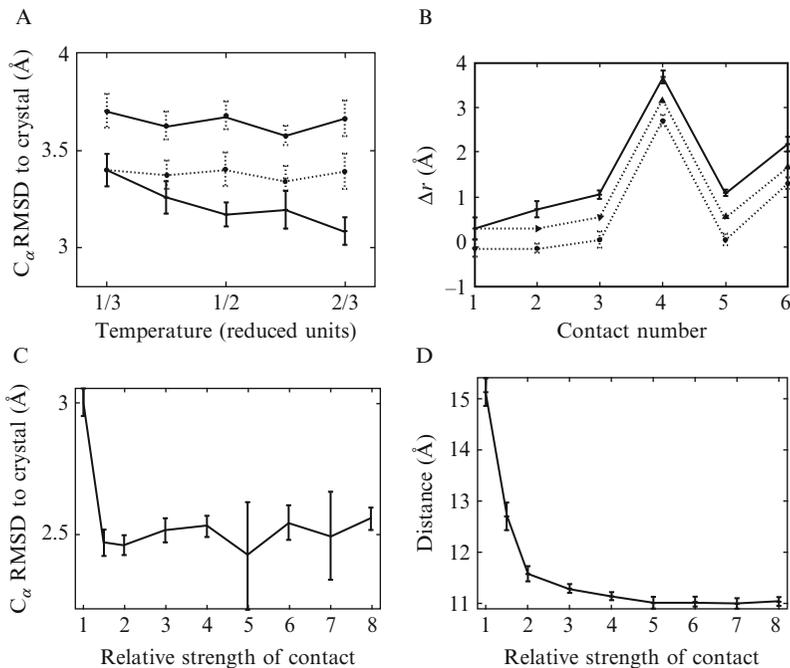


Figure 3.6 Robustness of parameters for the Spo0F/Spo0B system. Ten docking simulations using the DCA predicted interface contacts (full: contact strength 5 kT and 6 Å, striped: 5 kT and 6.5 Å and thinly striped: 5 kT and 7 Å) lead to comparably good results. (A) We find successful docking simulations (all contacts are formed in the docked conformation) for different temperatures with a RMSD difference to the crystal around 3–3.5 Å C_{α} -RMSD. (B) The differences Δr of the six contact distances with the crystal distance show comparable deviations (see also Table 3.1). While (A) suggests a contact distance of 6 Å optimal, (B) suggests 7 Å to be a slightly better choice. We arbitrarily choose 7 Å as default value for further simulations. (C, D) Due to the perfect conservation of the His-Asp pair in the two proteins, DCA in principle cannot detect them as an interactions (covariance requires variance). The prediction quality improves when including this contact additionally. The strength is relative to the DCA-contacts (5 kT , 7 Å) and the distance 11 Å allows to accommodate of the phosphoryl group (assumed to be roughly 4 Å large). It shows that this additional contact needs to be strong enough ($> 7.5 kT$) to compete with the DI contacts to improve the prediction quality.

For the docking simulations the temperature is kept constant by the Berendsen algorithm (Berendsen *et al.*, 1984) with a coupling constant of 1. Each docking simulation runs 2.5 Mio time steps of 0.0025 using the described center-of-mass force and subsequent 0.5 Mio time steps without the center-of-mass force, running for a total of roughly 20 h on a typical CPU (Spo0F/Spo0B system, 4164 atoms). Having tested various sets of parameters (see Fig. 3.6 and Table 3.1), we suggest $T = 2/3$ and a combination of a contact strength of $5kT$ with a contact distance of 7 \AA for the DCA-contacts and $15kT/11 \text{ \AA}$ for the His-Asp contact for simulations.

2.2. Relaxation in an empirical force field

It is possible to relax the docked complexes additionally in an empirical all-atom force field for refinement. Here, we use AmberF99 (Wang *et al.*, 2000) (<http://chemistry.csulb.edu/ffamber>) with explicit Tip3p solvent and counterions (Jorgensen *et al.*, 1983), a time step of 0.002 fs, and particle mesh Ewalds electrostatics (Essmann *et al.*, 1995). This refinement aims at removing artifacts from different physical environments for the isolated and docked proteins.

There are several pitfalls/common problems when starting such simulations. First, the input pdb-file must be modified to accommodate the possibility of charged amino acids. Here we treat all LYS as LYP, CYS as CYN, and HIS as HID. Also, the start and end of each chain has to be identified (e.g., ASP to CASP or NASP indicating it being on the C- or N-terminal). For these changes, it might be necessary to add/remove some atoms. This can be done manually or by using homology modeling software (Eswar *et al.*, 2008). After that, it is necessary to minimize the structure for later simulations. We find it useful to first minimize the structure, then add solvent molecules (commands: editconf, genbox), minimize again, add counterions (genion), and minimize again. Afterward the simulation can be started.

While we see some relaxation of the sidechains in this simulation, the backbone shows only minor movements. The resulting structures are in high agreement ($\text{RMSD} \leq 3 \text{ \AA}$ excluding the mobile C-termini) with highly similar contact maps to the complexed crystal structure (1f51) (see Figs. 3.4 and 3.5) (Schug *et al.*, 2009b).

3. SUMMARY

We described the detailed MAGMA method that exemplified the feasibility of integrating sequence-based genomic analysis with molecular simulation to generate structural models of a signal transduction complex

at a resolution matching experimental accuracy (Schug *et al.*, 2009b). DCA described in the previous chapter was shown to give sufficient information to successfully dock the Spo0B/Spo0F system. The parameters for the SBS simulations are robust toward slight variations without significant changes of the resulting structure. This allows tuning and refining them for new specific systems or questions. We are confident MAGMA will successfully introduce other TCS or, more general, short-lived complex structures ruled by transient interactions, and allow concurrent simulation of the conformational and functional motions of the complex, such as those during the autophosphorylation reaction or phosphoryl-transfer reaction.

ACKNOWLEDGMENTS

This work was supported by the Center for Theoretical Biological Physics (CTBP) sponsored by the NSF (Grant PHY-0822283) with additional support from NSF grant MCB-0543906 (J. N. O.), NIH grant R01GM077298 (T. H.), and by NIH grant R01GM019416 (J. A. H.).

REFERENCES

- Adcock, S. A., and McCammon, J. A. (2006). Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**, 1589–1615.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
- Best, R. B., Buchete, N. V., and Hummer, G. (2008). Are current molecular dynamics force fields too helical? *Biophys. J.* **95**, L07–L09.
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein-folding—A synthesis. *Proteins* **21**, 167–195.
- Burbulys, D., Trach, K. A., and Hoch, J. A. (1991). Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell* **64**, 545–552.
- Chavez, L. L., Onuchic, J. N., and Clementi, C. (2004). Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* **126**, 8426–8432.
- Cheung, M. S., Finke, J. M., Callahan, B., and Onuchic, J. N. (2003). Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J. Phys. Chem. B* **107**, 11193–11200.
- Clementi, C., and Plotkin, S. S. (2004). The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci.* **13**, 1750–1766.
- Clementi, C., Nymeyer, H., and Onuchic, J. N. (2000). Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953.

- Clementi, C., Jennings, P. A., and Onuchic, J. N. (2001). Prediction of folding mechanism for circular-permuted proteins. *J. Mol. Biol.* **311**, 879–890.
- Essmann, U., Perara, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle Ewald method. *J. Chem. Phys.* **103**, 8577.
- Eswar, N., Eramian, D., Webb, B., Shen, M.-Y., and Sali, A. (2008). Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426**, 145–159.
- Finke, J. M., Cheung, M. S., and Onuchic, J. N. (2004). A structural model of polyglutamine determined from a host-guest method combining experiments and landscape theory. *Biophys. J.* **87**, 1900–1918.
- Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science* **254**, 1598–1603.
- Fujitsuka, Y., Takada, S., Luthey-Schulten, Z. A., and Wolynes, P. G. (2004). Optimizing physical energy functions for protein folding. *Proteins* **54**, 88–103.
- Galperin, M. Y. (2006). Structural classification of bacterial response regulators: Diversity of output domains and domain combinations. *J. Bacteriol.* **188**, 4169–4182.
- Gambin, Y., Schug, A., Lemke, E. A., Lavinder, J. J., Ferreon, A. C., Magliery, T. J., Onuchic, J. N., and Deniz, A. A. (2009). Direct single-molecule observation of a protein living in two opposed native structures. *Proc. Natl. Acad. Sci. USA* **106**, 10153–10158.
- Hardin, C., Pogorelov, T. V., and Luthey-Schulten, Z. (2002). Ab initio protein structure prediction. *Curr. Opin. Struct. Biol.* **12**, 176–181.
- Hoch, J. A. (2000). Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**, 165–170.
- Hoch, J. A., and Varughese, K. I. (2001). Keeping signals straight in phosphorelay signal transduction. *J. Bacteriol.* **183**, 4941–4949.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**(33–8), 27–28.
- Janin, J., Henrick, K., Moult, J., Eyck, L. T., Sternberg, M. J., Vajda, S., Vakser, I., and Wodak, S. J. (2003). CAPRI: A critical assessment of predicted interactions. *Proteins* **52**, 2–9.
- Jiang, M., Shao, W., Perego, M., and Hoch, J. A. (2000). Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **38**, 535–542.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
- Kutzner, C., van der Spoel, D., Fechner, M., Lindahl, E., Schmitt, U. W., de Groot, B. L., and Grubmüller, H. (2007). Speeding up parallel GROMACS on high-latency networks. *J. Comput. Chem.* **28**, 2075–2084.
- Lammert, H., Schug, A., and Onuchic, J. N. (2009). Robustness and generalization of structure-based models for protein folding and function. *Proteins* **77**, 881–891.
- Levy, Y., and Onuchic, J. N. (2006). Mechanisms of protein assembly: Lessons from minimalist models. *Acc. Chem. Res.* **39**, 135–142.
- Levy, Y., Onuchic, J. N., and Wolynes, P. G. (2007). Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **129**, 738–739.
- Linhananta, A., and Zhou, Y. Q. (2002). The role of sidechain packing and native contact interactions in folding: Discontinuous molecular dynamics folding simulations of an all-atom G(o)ver-bar model of fragment B of Staphylococcal protein A. *J. Chem. Phys.* **117**, 8983–8995.
- Mascher, T., Helmann, J. D., and Uuden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol. Mol. Biol. Rev.* **70**, 910–938.

- Mukhopadhyay, D., Sen, U., Zapf, J., and Varughese, K. I. (2004). Metals in the sporulation phosphorelay: Manganese binding by the response regulator Spo0F. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 638–645.
- Oliveira, L. C., Schug, A., and Onuchic, J. N. (2008). Geometrical features of the protein folding mechanism are a robust property of the energy landscape: A detailed investigation of several reduced models. *J. Phys. Chem. B* **112**, 6131–6136.
- Onuchic, J. N., and Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
- Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997). Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600.
- Schug, A., and Wenzel, W. (2006). An evolutionary strategy for all-atom folding of the 60-amino-acid bacterial ribosomal protein L20. *Biophys. J.* **90**, 4273–4280.
- Schug, A., Fischer, B., Verma, A., Merlitz, H., Wenzel, W., and Schoen, G. (2005a). Biomolecular structure prediction stochastic optimization methods. *Adv. Eng. Mat.* **7**, 1005–1009.
- Schug, A., Herges, T., Verma, A., Lee, K. H., and Wenzel, W. (2005b). Comparison of Stochastic optimization methods for all-atom folding of the Trp-cage protein. *Chemphyschem* **6**, 2640–2646.
- Schug, A., Whitford, P. C., Levy, Y., and Onuchic, J. N. (2007). Mutations as trapdoors to two competing native conformations of the Rop-dimer. *Proc. Natl. Acad. Sci. USA* **104**, 17674–17679.
- Schug, A., Hyeon, C., and Onuchic, J. (2009a). Coarse-grained structure-based simulations of proteins and RNA. In “Coarse-Graining of Condensed Phase and Biomolecular Systems,” (G. A. Voth, ed.), pp. 123–140. CRC Press, Boca Raton, FL.
- Schug, A., Weigt, M., Onuchic, J. N., Hwa, T., and Szurmant, H. (2009b). High resolution complexes from integrating genomic information with molecular simulation. *Proc. Natl. Acad. Sci. USA* **106**, 22124–22129.
- Shimada, J., Kussell, E. L., and Shakhnovich, E. I. (2001). The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *J. Mol. Biol.* **308**, 79–95.
- Szurmant, H., White, R. A., and Hoch, J. A. (2007). Sensor complexes regulating two-component signal transduction. *Curr. Opin. Struct. Biol.* **17**, 706–715.
- Thirumalai, D., and Hyeon, C. (2005). RNA and protein folding: Common themes and variations. *Biochemistry* **44**, 4957–4970.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718.
- Varughese, K. I., Madhusudan, Zhou, X. Z., Whiteley, J. M., and Hoch, J. A. (1998). Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Mol. Cell.* **2**, 485–493.
- Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) perform in calculating conformational energies of organic and biological molecules? *J. Chem. Phys.* **21**, 1049–1074.
- Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. USA* **106**, 67–72.
- Whitford, P. C., Miyashita, O., Levy, Y., and Onuchic, J. N. (2007). Conformational transitions of adenylate kinase: Switching by cracking. *J. Mol. Biol.* **366**, 1661–1671.
- Whitford, P. C., Noel, J. K., Gosavi, S., Schug, A., Sanbonmatsu, K. Y., and Onuchic, J. N. (2009a). An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430–441.
- Whitford, P. C., Schug, A., Saunders, J., Hennelly, S. P., Onuchic, J. N., and Sanbonmatsu, K. Y. (2009b). Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function. *Biophys. J.* **96**, L7–L9.

- Zapf, J., Sen, U., Madhusudan, Hoch, J. A., and Varughese, K. I. (2000). A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure* **8**, 851–862.
- Zhou, Y., Zhang, C., Stell, G., and Wang, J. (2003). Temperature dependence of the distribution of the first passage time: Results from discontinuous molecular dynamics simulations of an all-atom model of the second beta-hairpin fragment of protein G. *J. Am. Chem. Soc.* **125**, 6300–6305.