The Role of Water in the Stability of Wild Type and Mutant Insulin Dimers

Shampa Raghunathan,[†] Krystel El. Hage,[†,‡] Jasmine L. Desmond,[†] Lixian Zhang,[†] and Markus Meuwly[*,†]

[†]*Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056, Basel, Switzerland*

[‡]*Current address: 1. Department of Fundamental Oncology, Lausanne University, Ludwig Institute for Cancer Research, Route de la Corniche 9A, 1066 Epalinges, Switzerland 2. SIB Swiss Institute of Bioinformatics, Quartier UNIL-Sorge, Batiment Genopode, 1015 Lausanne, Switzerland*

E-mail: m.meuwly@unibas.ch

Phone: +41 (0)61 207 38 21

Abstract

Insulin dimerization and aggregation play important roles in the endogenous delivery of the hormone. One of the important residues at the insulin dimer interface is Phe^{B24} which is an invariant aromatic anchor that packs towards its own monomer inside a hydrophobic cavity formed by Val^{B12}, Leu^{B15}, Tyr^{B16}, Cys^{B19} and Tyr^{B26}. Using molecular dynamics and free energy simulations in explicit solvent, the structural and dynamical consequences of mutations of Phe at position B24 to Gly, Ala, and d-Ala and the des-PheB25 variant are quantified. Consistent with experiments it is found that the Gly and Ala modifications lead to insulin dimers with reduced stability by 4 and 5 kcal/mol from thermodynamic integration and 4 and 8 kcal/mol from results

using MM-GBSA, respectively. Given the experimental difficulties to quantify the thermodynamic stability of modified insulin dimers, such computations provide a valuable complement. Interestingly, the Gly-mutant exists as a strongly and a weakly interacting dimer. Analysis of the molecular dynamics simulations shows that this can be explained by water molecules that replace direct monomer-monomer H-bonding contacts at the dimerization interface involving residues B24 to B26. It is concluded that such solvent molecules play an essential role and must be included in future insulin dimerization studies.

Introduction

Insulin is a small, aggregating protein that plays an eminent role in regulating glucose uptake in cells. In its crystal form, the WT hormone is a hexamer consisting of three dimers with either two or four Zn atoms bound to it.[1] Each dimer consists of two monomers (chain A with 21 amino acids and chain B with 30 amino acids), connected by two inter-chain (Cys^{A7}–Cys^{B7}, Cys^{A20}–Cys^{B19}) and one intra-chain (Cys^{A6}–Cys^{A11}) disulfide bonds (see Figure 1). Under physiological conditions insulin monomers readily aggregate to form dimers. Experimental and computational studies have found that the main stabilizing contributions to self-association are nonpolar interactions through directionality provided by hydrogen bonds.[1–5] Dissociation of the dimer to form two monomeric insulins is of great physiological importance as the monomer is the functionally relevant state of the hormone. One suggested possibility to suppress aggregation is to modify the dimerization interface through suitable substitutions[3] or chemical modifications.[6]

The current view of the insulin structure-function relationship is derived primarily from insulin hexamer and dimer crystal structures, as well as from studies of the structure-activity correlations of chemically modified and/or naturally occurring mutant insulins in solution.[7–15] Mutagenesis of the dimer-forming surface of insulin can yield analogues with a

reduced tendency to aggregate and pronounced differences in the pharmacokinetic properties with potentially promising therapeutic applications.[7,16] Typical experimental methods for quantitative studies of insulin dimerization are isothermal titration calorimetry (ITC)[3] or NMR spectroscopy. ITC requires relatively high protein concentrations, while NMR spectroscopy can be slow for such purposes. Several NMR studies of active monomeric insulin mutants show a rearrangement of the C-terminal end of chain B.[8,14] The removal of residues B26–B30 in despentapeptide prevents dimerization without any significant changes in the rest of the molecule.[1,3,17] This truncated insulin is at least 50 % as potent as human insulin.[3,15] A combined Raman spectroscopy and microscopy study of insulin in different aggregation states (monomer, dimer, hexamer and fibril) shows that dimerization damps fluctuations at an intermolecular β-sheet.[18] Experimental alanine scanning finds substitution of alanine at various positions to reduce insulin affinity for the receptor by more than 20-fold.[19] While the residues that are most likely to be directly involved in binding are A1, A2, A3, A19, B12, B23 and B24, any substitution of residues A1–A3 has been shown to impair function.[15]

Phenylalanine at position B24 is invariant among insulin sequences and is located at the dimerization interface maintaining the orientation of the B-chain of the monomer.[8,20–22] These observations together with studies of low-potency B24 analogues suggest that the Phe^{B24} amino acid residue plays an important role[23] in the activity of insulin, while Ser^{B24} (Ref.[24,25]), Leu^{B24} (Ref.[26]), and His^{B24} (Ref.[27]) analogues show reduced binding potency. However, it was also found that certain B24 substitutions, such as Gly,[16,28] D-Ala,[11,29] D-His,[27] Met and Cha[30] are well tolerated in view of the affinity of insulin to its receptor. The bioactivity of such insulins has also been described as "anomalous" because it can not be readily explained by crystal models.

Experimental data for dimerization free energies of insulin analogues is scarce due to several experimental challenges. The role of Phe^{B24} in stabilizing the insulin dimer has also

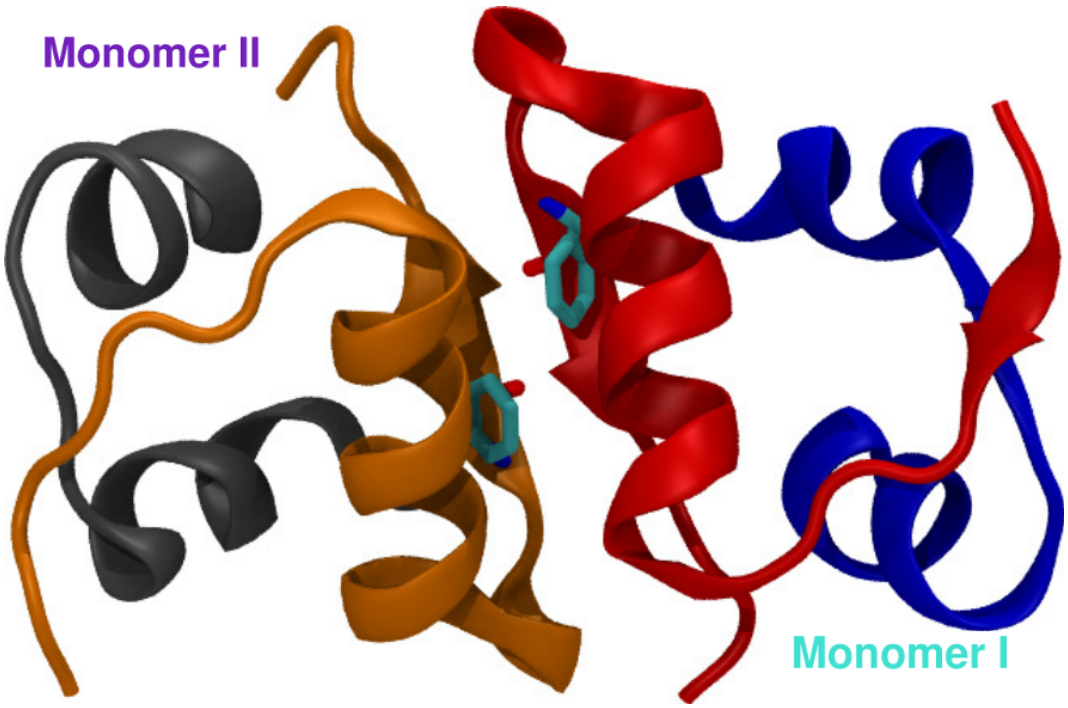

Figure 1: Structure of insulin dimer (PDB Code: 4INS). Chain A1 (blue), Chain B1 (red), Chain A2 (grey), Chain B2 (orange). Residue 24 in both monomers is shown in stick representation.

been studied to some extent. Unlike WT insulin, the Gly^{B24} mutant does not dimerize in aqueous solution at pH 1.9.[8] Furthermore, alanine scanning of the dimerization interface revealed that the Ala^{B24} analogue is monomeric and does not readily aggregate.[31,32] This suggests that the Ala^{B24} and Gly^{B24} analogue dimers are less stable than the WT dimer. Furthermore, ITC measurements of N-methylated insulin dimer analogues at positions B24, B25, and B26[33,34] revealed considerably reduced (by a factor of 5) dimerization capabilities compared with human insulin.

Because the dimer↔monomer equilibrium is one of the essential steps in forming the receptor binding-competent monomeric form of insulin and the process is difficult to study quantitatively by experiments, MD simulations are an attractive alternative to characterize the stability of insulin dimers. In the present work the stability and dynamics of the insulin dimer analogues (Gly^{B24}, Ala^{B24}, D-Ala^{B24} and des-Phe^{B25}) are investigated using atomistic

simulations in explicit solvent. The relative stabilities of the analogues are compared with the native WT dimer and with qualitative results from experiments. Both, protein-protein and protein-water contacts are analyzed to characterize the role of water[35] and hydrogen bonding at the dimer-forming surface. Previous computational studies of the WT dimer[4,36] and of alanine scans[37] have provided important complementary information to experimental characterizations and are a rational basis to extend such an approach to modified insulins using explicit free energy simulations. Hence, the present work lays the foundation to extend insulin dimerization studies to arbitrary modifications at the interface.

This article is organized as follows: Section 2 introduces the methods. In Section 3 the dynamics and stability of insulin dimers and the inter-molecular hydrogen bonding between monomer-monomer and water-protein (water-bridged) is presented and discussed. Finally conclusions are drawn in Section 4.

Methods

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were carried out using CHARMM[38] together with the "all-atom" CHARMM22[39] force field including the CMAP correction[40,41] and periodic boundary conditions (PBC). Additional validation simulations for the insulin dimer were carried out using Gromacs[42] and the CHARMM36 force field[43] as described in the supporting information. The starting coordinates for the MD simulations were the X-ray structure of the WT porcine insulin dimer resolved at 1.5 Å (Protein Data Bank (PDB[44,45]), Code: 4INS.[1] The structure contains the coordinates of the insulin dimer and two aggregated Zn atoms. Zn atoms are removed as they been shown to be relevant only for hexamer formation. Hydrogen atoms were added to the X-ray structure. The resulting structure was used to gen-

erate mutants computationally. For this, the Phe residue at position B24 was mutated into Glycine (Gly), Alanine (Ala), and D-Alanine (D-Ala) yielding mutants Gly^{B24}, Ala^{B24}, and D-Ala^{B24}, respectively. Furthermore, the wildtype (WT) insulin dimer without the Phe^{B25} amino acid on both monomers, a des-Phe^{B25} mutant was also studied.

The wildtype dimer and mutants were solvated in a 77.6 $\times$ 62.8 $\times$ 55.8 Å box of TIP3P[46] water molecules. All MD simulations were carried out in explicit water. Water molecules overlapping the protein were removed which leads to a system with approximately 1550 protein atoms and 8495 water molecules. The solvent was equilibrated at 300 K for 30 ps with the insulin frozen. Then 2000 steps of steepest descent (SD) minimization were carried out. The entire system was heated to 300 K during 15 ps using harmonic constraints with a force constant of 5 kcal/mol $Å^2$ on the position of the backbone atoms. The system was further equilibrated for 120 ps by gradually decreasing harmonic constraints on the backbone atoms. For all simulations the Verlet leapfrog integrator was used for time propagation with a time step of 1 fs. A 12 Å cutoff was applied to the shifted electrostatic and switched van der Waals interactions and images for periodic boundary conditions were updated every 10 time steps. All distances to hydrogen atoms were constrained by using SHAKE.[47] For the WT dimer and mutants (Gly^{B24}, Ala^{B24}, D-Ala^{B24} and des-Phe^{B25}) multiple individual trajectories were run starting from different structures taken from the equilibration run (Table S1). Simulations in the NPT and NVT ensembles were run using the extended system constant pressure and temperature (CPT) algorithms[48] with a Hoover thermostat.[49] In addition, microcanonical NVE simulations were run as well.

Calculation of Binding Free Energy

In the present work, the stability of the WT and mutant insulin dimers is determined from two complementary approaches. One of them is the molecular mechanics-generalized Born

surface area (MM-GBSA) approach[50,51] and the other one is thermodynamic integration.[52,53] The thermodynamic cycles used for computing the binding free energy ΔG_{bind} using MM-GBSA and thermodynamic integration (TI) are shown in Figures 2A and B, respectively. Computational details for MM-GBSA are provided in the Supporting Information.

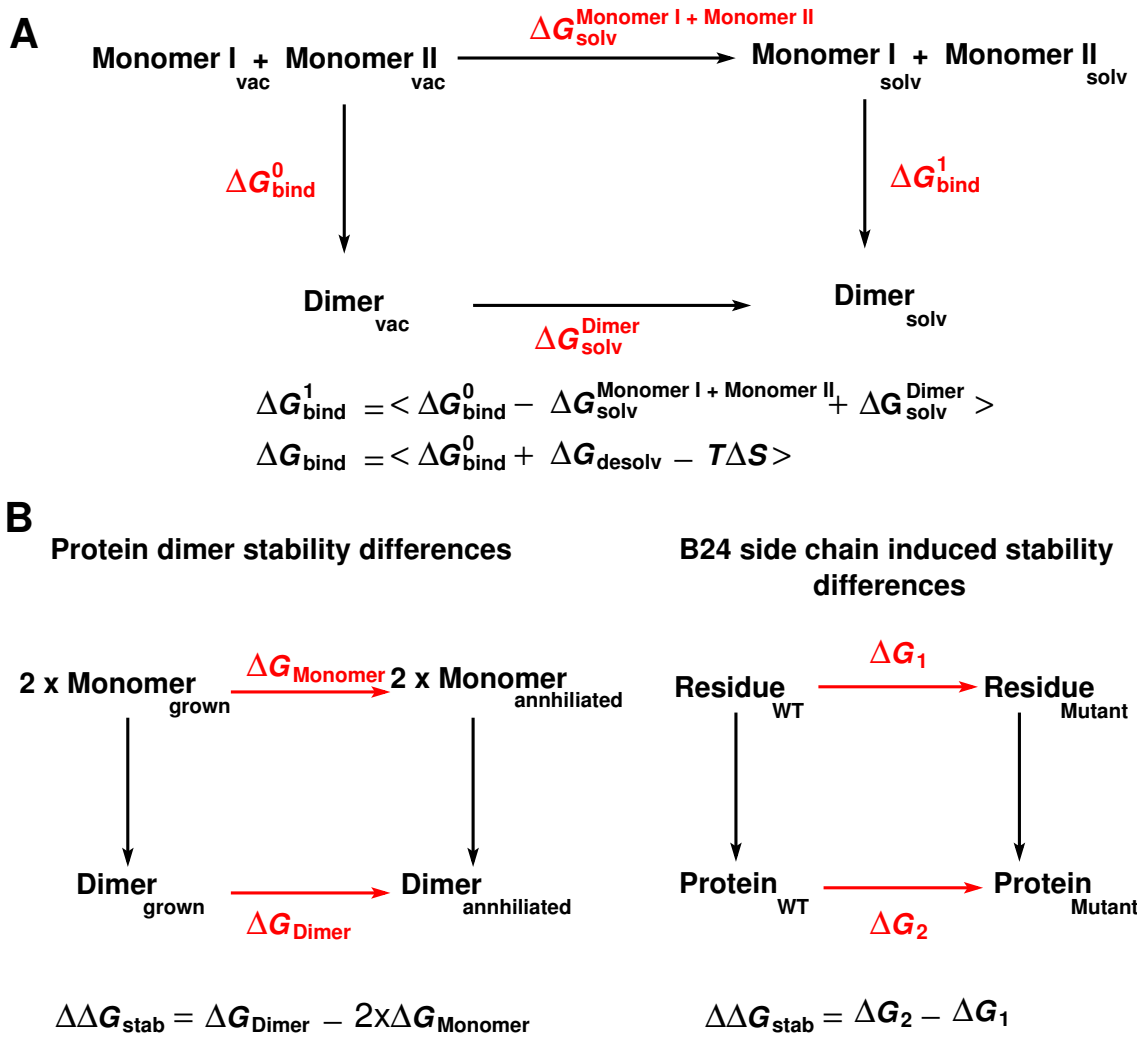

Figure 2: Thermodynamic cycles used to determine stabilization free energies from MM-GBSA simulations (A) and from thermodynamic integration (B). In (B, left panel) the thermodynamic cycle used to compute the protein dimerisation stability differences, and the right panel shows the thermodynamic cycle used to compute the stability differences by mutating the B24 side chain only.

To corroborate the results from the MM-GBSA simulations and to investigate the role of

water molecules on the thermodynamic stability, dimer stabilization free energies were also determined from thermodynamic integration (TI) in the presence of explicit water molecules. TI applies a scaling parameter λ to switch between an initial (state A, $\lambda = 0$) and the final (state B, $\lambda = 1$) state by gradually damping all nonbonded interactions. The $\lambda = 0$ and $\lambda = 1$ states correspond to the grown and annihilated nonbonded interactions on either the protein dimer or on the monomer, respectively. Initial coordinates were taken from the equilibrated simulations. TI simulations were performed using CHARMM's PERT module using soft-core potentials[54–56] for the LJ interactions that applied only on the repulsive part of the LJ potential (it was used when the electrostatic interactions were turned off). For this computational approach, restraining potentials[57] affecting the translational, rotational and conformational freedom of the protein may be activated and released during the simulations to aid convergence and improve the sampling.

Free energy simulations at each $\lambda-$value were carried out in the NPT ensemble, using the Hoover heat-bath method[49] with pressure coupling at $T = 298$ K, $p = 1$ atm, and with the masses of the temperature and pressure piston set to roughly 20 % and 2 % of the system's mass, respectively. A friction coefficient of 50 ps^{-1} was used. The interval $0 < \lambda < 1$ was divided into 40 equidistant steps to ensure accuracy. For each $\lambda-$value the system was re-equilibrated for 60 ps followed by 100 ps of dynamics during which information was accumulated. λ was changed from initial to final value using the slow-growth protocol,[58] which allowed the system to re-equilibrate between steps. Contrary to the MM-GBSA simulations (see SI), TI run in this fashion includes all enthalpic and entropic contributions. Assuming that the change in the entropic contribution $\Delta\Delta S_{\mathrm{WT-Mutant}}$ remains approximately constant for the various mutants it is expected that the *stability ranking* from the two methods remains the same which is a testable hypothesis.

The protein dimer stability difference $\Delta\Delta G_{\mathrm{stab}} = \Delta G_{\mathrm{dimer}} - 2 \times \Delta G_{\mathrm{monomer}}$ was computed

within the "same trajectory method"[4] such as to close the thermodynamic cycle, see Figure 2B (left panel).

The free energy of mutating Phe^{B24} (WT) into Gly and Ala was also calculated directly using the dual topology approach[54,59] for mutations F24G and F24A in the dimer (Figure 2, Panel B right). The effect of mutating the B24 side chain on the stability of the protein $\Delta\Delta G$ was calculated according to ΔG_2 - ΔG_1, where ΔG_1 and ΔG_2 are the free energies of mutating the side chain of a WT residue into another residue (mutant), in the aqueous phase, as a sole residue and in the protein, respectively. For these simulations, no restraints were used and the interval $0 < \lambda < 1$ was divided into 34 steps. Windows at the two ends of the λ interval were more finely spaced. For each of these steps the system was re-equilibrated for 30 ps followed by 60 ps of dynamics during which information was accumulated. The results are averages and standard deviations of five runs.

Results and Discussion

In the following, first, the stabilities from MM-GBSA and thermodynamics integration simulations of the dimers are discussed. Then, the findings are discussed in the context of the dynamics and interactions along the dimerization interface. Finally, the role of water molecules is considered.

Dimerization Free Energies

One of the main objectives of the present study is to determine the relative stabilities of mutated insulin dimer analogues relative to the WT dimer. Experimentally, a dimerization energy of -7.2 ± 0.8 kcal/mol[60] in favour of the dimer was determined for WT insulin

which compares with $\Delta G_{\mathrm{bind}} = -11.9 \pm 6.7$ kcal/mol (absolute binding free energy) and $\langle \Delta G^{0}_{\mathrm{bind}} \rangle + \langle \Delta G_{\mathrm{desolv}} \rangle = -38.7 \pm 5.8$ kcal/mol (enthalpic contribution to ΔG_{bind}) from previous molecular dynamics simulations.[4]

Table 1 reports the different contributions to the calculated absolute binding free energies of dimerization for the different insulin analogues including the WT. The enthalpic contribution $G_{\mathrm{enthalpic}}$ is the sum of E_{vdW}, E_{elec}, $G_{\mathrm{solv,elec}}$, and $G_{\mathrm{solv,nb}}$ which is ~ -45 kcal/mol for the WT dimer. Compared to this, Gly$^{\mathrm{B24}}$, Ala$^{\mathrm{B24}}$, and D-Ala$^{\mathrm{B24}}$ (approximately -38, -38, -33 kcal/mol, respectively) are enthalpically less stable. The des-Phe$^{\mathrm{B25}}$ analogue is least stable (~ -16 kcal/mol). The destabilization energy of about 7–12 kcal/mol for the different energetically low-lying analogues of the insulin dimer is related to replacing the two Phe$^{\mathrm{B24}}$ residues (in both monomers) which contribute an average stabilization of -3.96 and -3.36 kcal/mol, to the total stabilization of the WT protein (see Figure S1). These contributions are consistent with results reported in an earlier MM-GBSA study which found -3.92 and -2.68 kcal/mol, respectively.[4] Experimentally, the des-Phe$^{\mathrm{B25}}$ insulin dimer was also found to be unstable as it reported exclusively monomeric insulin for this variant.[2] The individual contributions (see Table 1) to the binding free energy suggest that stabilization is predominantly due to nonpolar terms E_{vdW} and $G_{\mathrm{solv,nb}}$. The favorable E_{elec} contribution from the two monomers is canceled by the desolvation energy $G_{\mathrm{solv,elec}}$ upon dimerization. This is found for all modified insulins investigated here and supports previous investigations of the WT.[4]

Including the entropic contribution $T(\Delta S_{\mathrm{trans}} + \Delta S_{\mathrm{rot}} + \Delta S_{\mathrm{vis}})$ to ΔG_{bind}, allows more direct comparison with experimentally determined values. For the WT the stabilization free energy is $\Delta G_{\mathrm{bind}} = -16.0 \pm 6.9$ kcal/mol when accounting for entropic contributions, in qualitative agreement with the experimental binding free energy of -7.2 ± 0.8 kcal/mol.[60] As a comparison, previous simulation work based on one trajectory found an enthalpic stabilization

Table 1: Binding free energy (kcal/mol) decomposition for the insulin dimerization of various insulin analogues: vdW, electrostatic, and solvation (elec. and nonpolar) contributions, using the same trajectory method by MD simulations.

	WT	Gly^{B24}	Ala^{B24}	D-Ala^{B24}	des-Phe^{B25}
$\langle \Delta E_{\mathrm{vdw}} \rangle$	-66.59(4.4)	-53.39(4.3)	-71.35(4.0)	-60.39(5.1)	-44.47(5.2)
$\langle \Delta E_{\mathrm{ele}} \rangle$	-114.38(34.3)	-107.84(40.7)	-111.55(26.5)	-104.85(33.9)	-56.15(32.2)
$\langle \Delta G_{\mathrm{ele,desolv}} \rangle$	146.23(32.6)	132.83(38.4)	155.99(23.8)	142.77(30.8)	92.43(29.6)
$\Delta G_{\mathrm{np,desolv}} \rangle$	-10.28(0.6)	-9.76(0.5)	-11.49(0.4)	-10.54(0.5)	7.99(0.6)
$\langle \Delta G^0_{\mathrm{bind}} \rangle +$ $\langle \Delta G_{\mathrm{desolv}} \rangle$	-45.02(6.5)	-38.16(7.3)	-38.40(7.5)	-33.00(9.5)	-16.19(6.5)
$-T\langle \Delta S_{\mathrm{trans}} \rangle$	13.02[a]	13.02[a]	13.02[a]	13.02[a]	13.00[a]
$-T\langle \Delta S_{\mathrm{rot}} \rangle$	14.36(0.02)	14.31(0.02)	14.35(0.01)	14.33(0.01)	14.32(0.02)
$-T\langle \Delta S_{\mathrm{vib}} \rangle$	1.61(3.7)	-1.43(4.0)	3.50(3.5)	-1.87(4.1)	-4.52(3.6)
$-T\langle \Delta S \rangle$	28.99(3.7)	25.90(4.0)	30.86(3.5)	25.47(4.1)	22.80(3.6)
$\langle \Delta G_{\mathrm{bind}} \rangle$	-16.03(6.9)	-12.26(8.0)	-7.54(8.0)	-7.53(10.9)	6.61(7.2)

[a] The standard deviation of $-T\langle \Delta S_{\mathrm{trans}} \rangle$ is not defined, because it is a function of mass, which has constant value.
† The standard state is taken to be 1 M as was used by Tidor *et al.*[61]
‡ Symmetry corrected using $\sigma = 2$ in equation 13 from Ref.[61]

of -38.6 ± 5.8 kcal/mol for the WT dimer which decreased to -11.9 ± 6.7 kcal/mol when including entropic contributions.[4] This is consistent with earlier studies[61] on the dimerization of WT insulin which found an unfavourable entropic contribution of ≈ 30 kcal/mol (depending on the size and shape of the protein). On the other hand, for *relative* stabilization free energies upon mutation the entropic part is less important except for the contribution due to vibrations. For the protein variants considered here, $T\Delta S_{\mathrm{vib}}$ ranges from -2 to 3.5 kcal/mol and hence can contribute up to 6 kcal/mol to the differential stabilization of one protein variant relative to another one.

The results in Table 1 show that Ala^{B24} is the entropically least favored substitution while des-Phe^{B25} is most favoured. Adding the entropic and enthalpic contributions, the total binding free energies of dimerization of insulins leads to stabilisation ranging from -16 to -7 kcal/mol among the low-lying analogues, *i.e.*, WT, Gly^{B24}, Ala^{B24} and D-Ala^{B24}. The

des-Phe^{B25} mutant with ΔG_{bind} of ~ 6 kcal/mol is energetically unfavorable and is expected to be monomeric in solution.

Overall, 46 independent (10 for WT, Gly^{B24}, Ala^{B24} and des-Phe^{B25}, 6 for D-Ala^{B24}) free energy simulations were performed, each 10 ns long, which amounts to a typical aggregate of 100 ns for each system studied. These simulations were also run in different statistical mechanical ensembles and Table S1 in the SI provides a comprehensive summary. Considering the enthalpic part of ΔG_{bind} for all the trajectories (WT and B24 insulin dimer analogues) two situations can be distinguished (see Table S1, and Figures S2 to S7 in the SI for illustrations): 1) Stable, low-lying (~ -45 kcal/mol) analogues with binding energies comparable to the WT dimer, and 2) Less-stable, high-lying (~ -20 kcal/mol) analogues with decreased stability. The ten different WT trajectories show similar binding energies. Even Ala^{B24}, and des-Phe^{B25} show similar binding energies among various trajectories. However, Gly^{B24} and D-Ala^{B24} show appreciable differences in binding energies among various trajectories. Specifically, for the Gly^{B24} mutant a strongly (SI) and a weakly interacting (WI) dimer is found which will be discussed further below.

Dimer stabilization free energies were also determined from thermodynamic integration (TI, see methods) which provide a direct validation of the MM-GBSA results. TI simulations were carried out for the WT (Phe^{B24}) and the 3 mutants (Gly^{B24}, Ala^{B24} and des-Phe^{B25}) using the single topology method. The computed stabilization free energy of -8.4 kcal/mol for the WT dimer differs by about 1 kcal/mol from the experimentally measured value of -7.2 ± 0.8 kcal/mol[60] which provides a direct validation of the TI simulations. Relative to WT, the Gly^{B24} and Ala^{B24} mutants are destabilized by $\Delta\Delta G^{\text{TI}}_{\text{WT/GlyB24}} = 3.8$ kcal/mol and $\Delta\Delta G^{\text{TI}}_{\text{WT/AlaB24}} = 5.4$ kcal/mol, respectively, compared with $\Delta\Delta G^{\text{MM-GBSA}}_{\text{WT/GlyB24}} = 3.7$ kcal/mol and $\Delta\Delta G^{\text{MM-GBSA}}_{\text{WT/AlaB24}} = 8.5$ kcal/mol from MM-GBSA simulations, see Table 1. The des-Phe^{B25} variant is unstable in TI as was also found from MM-GBSA. Hence, the TI and

MM-GBSA simulations are consistent with one another and support the experimental observation that the Ala^{B24} mutant is marginally stable/unstable in solution[31,32] which serves as an additional validation of the present simulations.

Table 2: Stability free energy (kcal/mol) of the various insulin dimer analogues. The error on the computed values is reported in parentheses.

	WT	Gly^{B24}	Ala^{B24}	des-Phe^{B25}
$\Delta\Delta G_{\text{Stability}}$	−8.4 (0.2)	−4.6 (0.2)	−3.0 (0.2)	2.3 (0.2)

The free energy of mutating Phe^{B24} (WT) into Gly and Ala was also calculated directly (Figure 2, Panel B right) using the dual topology approach[54,59] for mutations F24G and F24A in the dimer. These computations yield $\Delta\Delta G_{\text{PheB24}\rightarrow\text{GlyB24}} = 3.2 \pm 0.2$ kcal/mol and $\Delta\Delta G_{\text{PheB24}\rightarrow\text{AlaB24}} = 4.2 \pm 0.2$ kcal/mol. In other words, the Gly^{B24} and Ala^{B24} mutants are destabilized by 3.2 and 4.2 kcal/mol relative to the WT protein, respectively, which compares and is consistent with values of 3.8 and 5.4 kcal/mol from the first set of TI simulations, see Table 2.

In summary, the MM-GBSA and TI simulations all agree in that WT (Phe^{B24}) is most stable, followed by Gly^{B24} and Ala^{B24} mutants. The des-Phe^{B25} variant is unstable. Furthermore, the differential stabilization free energies of two TI simulations differ by 1 to 1.5 kcal/mol which is, however, acceptable given the very different ways in which they were carried out.

The Weakly and Strongly Interacting Gly^{B24} Dimer

The MM-GBSA simulations found a strongly (SI) and a weakly (WI) interacting Gly^{B24} dimer. For the SI variant ($\Delta G \approx -48$ kcal/mol) the dimer is stabilized almost as strongly as the WT dimer, and for WI, the dimer is considerably destabilized by almost 30 kcal/mol

($\Delta G \approx -20$ kcal/mol). Figure S1 shows per-residue contributions to the total binding free energies of WT, SI and WI Gly^{B24} dimers. This analysis indicates that the differences mainly arise from contributions to the electrostatic $< E_{\text{elec}} >$ and solvation energy $< G_{\text{elec,desolv}} >$ (see Figure S2). For the WT and Gly^{B24}-SI insulin dimer the per-residue binding free energies follow a similar pattern whereas for Gly^{B24}-WI they differ (see Figure S1). For instance, most of the residues have favorable contributions to the total binding free energies for WT and SI (orange and yellow bars in Figure S1, respectively), whereas for the WI dimer (purple bars in Figure S1) these contributions are clearly reduced or even reversed which gives rise to reduced stabilization. Four residues of WI destabilize the dimer by > 2 kcal/mol.

One residue that contributes significantly to the differences between WI and SI is Glu^{B13} (see Figure S1) which makes unfavorable contributions of about 2.5 and 3.4 kcal/mol to the B1 and B2 chain, respectively. The differences were further analyzed and the electrostatic part was found to be primarily responsible for that, see Figure S8. The pattern of the per-residue contributions to the total dimerization free energy (see Figure S1) suggests that the two monomers in WT, SI and WI dimers are equivalent although they are not strictly symmetric as was reported, *e.g.*, for the crystal structure of the B9 (Ser→Glu) mutant insulin dimer[62] which was not symmetric.

In Figure 3B and D two H-bond distances between the side chains of residues His^{B10} and Glu^{B13} are reported for SI and WI, respectively. For Gly^{B24}-SI only a transiently formed intramonomer hydrogen bond is found, see Figure 3B. Contrary to that the two side chains His^{B10} and Glu^{B13} form an intramonomer H-bond which makes the donor and acceptor atoms of the two side chains unavailable for dimerization contacts, see Figure 3 D. This, in turn, reduces the stability as determined in the current protocol (MM-GBSA) and partially explains the difference of 20 kcal/mol between the SI and WI-dimer for the Gly^{B24} mutant. In addition to the loss of ≈ 6 kcal/mol, the dimer stability is also reduced due to four β-sheet

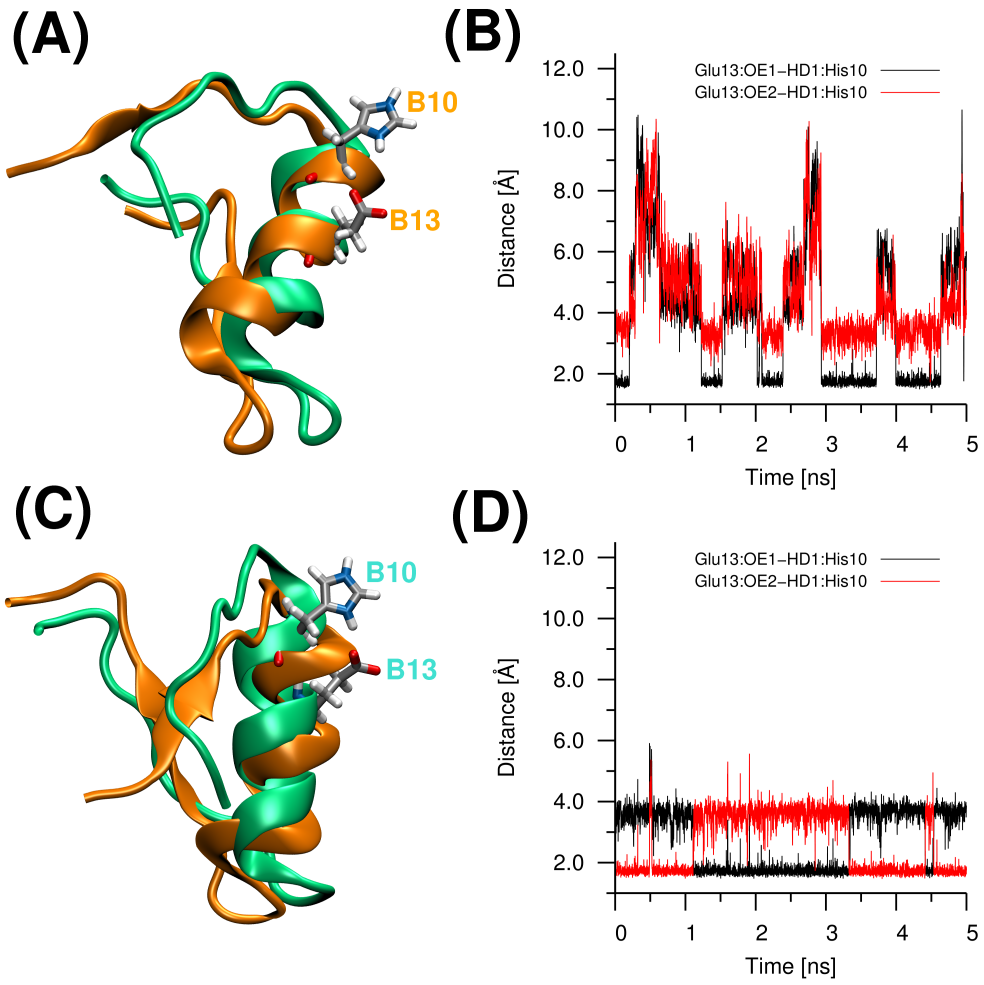

Figure 3: Superimposed (backbone) B1-chains of two snapshots for SI and WI variants of the Gly^{B24} insulin dimer analogue. The orange and green traces in (A) and (C) correspond to one monomer of the SI and WI dimer together with a CPK representation of the His^{B10} and Glu^{B13} residues. Panels (B) and (D) highlight the absence (B) and presence (D) of the hydrogen bond between His^{B10} and Glu^{B13}.

H-bonds which are almost absent in the WI dimer (average of 0.7 H-bonds) but exist most of the time for the SI (3.7 H-bonds), see Figure 4A.

Hydrogen Contacts at the Interface

In the insulin dimer the Phe^{B24}-Phe^{B25}-Tyr^{B26} segment of monomer I (chain B1) forms an antiparallel β-sheet with the adjacent Tyr^{B26}-Phe^{B25}-Phe^{B24} segment of monomer II (chain

B2, see Figure 4B). Inter-chain H-bonds are formed between Phe^{B24}(I) and Tyr^{B26}(II) and between Phe^{B24}(II) and Tyr^{B26}(I). Therefore, substitutions in this region influence dimer formation. Previous work[60] has reported that insulin dimerization is enthalpically controlled and the four inter-monomer H-bonds in the apolar environment are the prime driving force for insulin assembly. However, MD simulations[4] have shown that insulin dimerization primarily results from nonpolar interactions, in particular B24–B26 residues make the largest favorable contributions and the role of the H-bonds is to provide the necessary directionality of the interactions.

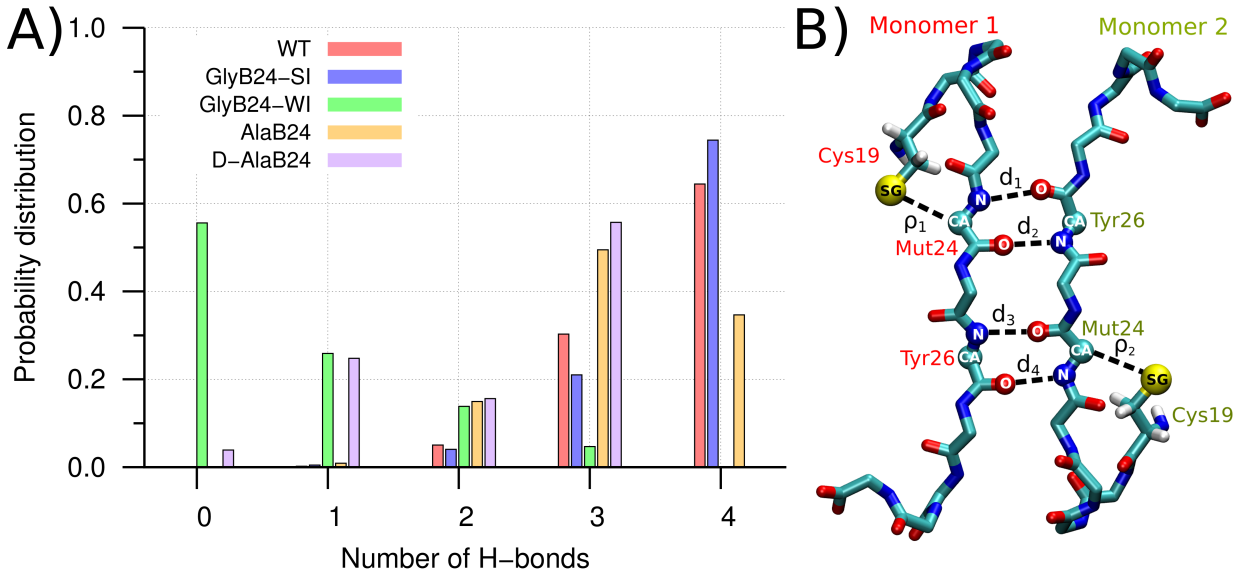

Figure 4: (A) The number of interfacial H-bonds between protein residues 24–26 of chains B1 and B2 and their probabilities for WT, Gly^{B24}-SI, Gly^{B24}-WI, Ala^{B24} and D-Ala^{B24} mutants. Average number of H-bonds are: 3.6 (WT), 3.7 (Gly^{B24}-SI), 0.7 (Gly^{B24}-WI), 3.2 (Ala^{B24}) and 2.2 (D-Ala^{B24}). Here, for analysis purposes, an H-bond is defined by a donor-acceptor (H-O) distance of $\leq$ 2.4 Å although several other complementary characterizations and criteria exist.[63] For these directional H-bonds at the dimerization interface the definition of an angle is not mandatory as the distinction between a "formed/established" and a "broken" H-bond for the purpose of the present analysis is straightforward, see Figure 6. Depending on this definition the persistence times of the dimerization contacts change somewhat but not the conclusions that are drawn from the analysis. (B) Backbone representation of the H-bonds d_1 to d_4 and the weaker CH–S contacts ρ_1 and ρ_2 along the dimerization interface, see also Figure 6.

Figure 4A reports the population of H-bonds between the two monomers involving residues 24–26 of chains B1 and B2. These involve (see Figure 4B) intermonomer contacts ρ_1 ($^{\mathrm{CysB19}}$SG–CA$^{\mathrm{XB24}}$) in monomer I, ρ_2 ($^{\mathrm{CysB19}}$SG–CA$^{\mathrm{XB24}}$) in monomer II, and intramonomer (dimerization) H-bonds d_1 ($^{\mathrm{XB24}}$N–O$^{\mathrm{TyrB26}}$), d_2 ($^{\mathrm{XB24}}$O–N$^{\mathrm{TyrB26}}$), d_3 ($^{\mathrm{TyrB26}}$N–O$^{\mathrm{XB24}}$) and d_4 ($^{\mathrm{TyrB26}}$O–N$^{\mathrm{XB24}}$), where X is Phe for WT or Ala and Gly for the mutants considered and where the first residue belongs to monomer I and the second corresponds to monomer II. Probability distributions of H-bonds (see Figure 4A) for WT and Gly$^{\mathrm{B24}}$-SI mutants are quite similar and show an average of 3.6 and 3.7 H-bonds, respectively, compared to only 0.7 H-bonds for the Gly$^{\mathrm{B24}}$-WI mutant. The H-bond distribution for the Ala$^{\mathrm{B24}}$ mutant (yellow) resembles that of the WT dimer with an average H-bond population of 3.2. For the D-Ala$^{\mathrm{B24}}$ analogue at most three H-bonds are found. Finally, no H-bond is present for the des-Phe$^{\mathrm{B25}}$ mutant indicating its inability to aggregate to a dimer.

The individual H-bonds at the dimerization interface can have different populations, see Table 3. For WT and Gly$^{\mathrm{B24}}$-SI the four inter-monomer H-bonds are present for $\sim 90\%$ of the simulation time. Specifically, for the WT occupations of 98 % were found for the $^{\mathrm{TyrB26}}$H–O$^{\mathrm{PheB24}}$ contacts and somewhat reduced occupancies for the $^{\mathrm{PheB24}}$H–O$^{\mathrm{TyrB26}}$ contacts (84 % for the B1–B2 and 78 % for the B2–B1 contact), see Table 3. Except for the B2–B1 $^{\mathrm{PheB24}}$H–O$^{\mathrm{TyrB26}}$ contact (47 %)[36] these occupancies are in reasonable agreement with previous work. Given the symmetric nature of the interface such a low occupancy is unexpected and the present results appear to be more realistic. Furthermore, the simulation time in the current work is considerably longer and the size of the simulation box is also larger, hence certain differences are not unreasonable.

For the weakly interacting dimer the H-bond occupancies are considerably reduced to between 0 and 34 %, see Table 3. These direct, inter-monomer H-bond occupancies correspond to the average number of H-bonds found in the H-bond distributions in Figure 4A. The

CH$\cdots$S contacts (which are not labelled "H-bonds" here) are weak interactions and for model systems such as $C_2H_2\cdots SH_2$ they were found to be stabilized by -1.34 kcal/mol at the MP2/aug-cc-pVQZ level of theory.[64]

Role of Water at the Interface

In search for a molecular explanation why the Gly^{B24} mutant is prone to destabilize, the hydration environment of the intermonomer H-bonds at the dimerization interface was further analyzed. The stability of the insulin dimer interface would decrease if monomer-monomer contacts would be replaced by protein-water interactions. Hence, the interfacial contacts, in particular the four β-sheet H-bonds (see Figure 4B) conserved and/or replaced by water-mediation in various insulin dimer analogues are explored. For this, the H-bonds around position B24 were analyzed and the results are summarized in Table 3.

For the WT dimer the water-protein H-bonds are occupied by 10 % to 30 % whereas for the strongly interacting Gly mutant almost no H-bonds to the water are found. Conversely, the weakly interacting Gly-dimer has occupations ranging from 15 % to 68 % which correlates with the reduced number of direct protein-protein H-bonds. This suggests that water molecules along the dimerization interface can replace direct protein-protein contacts. In general, low occupancy of protein-protein contacts implies a high population of water-mediated contacts.

In order to further characterize the stability, inter- and intramonomer contacts and close encounters with solvent water molecules for the different insulin dimers, and to support the findings described so far, additional 40 ns MD simulations for the Ala^{B24} and Gly^{B24} mutants and the WT were carried out, see Figure 6). The average RMSD compared to the 4INS reference structures are 2.0 Å, 2.5 Å, and 2.8 Å for the WT Ala^{B24}, and Gly^{B24} mutants,

Table 3: Monomer-monomer interfacial H-bonds involving residues 24 to 26 for WT, Gly^{B24}-SI and Gly^{B24}-WI, Ala^{B24} and D-Ala^{B24} mutants. The percentage of simulation time, during which the H-bonds were formed are reported for direct inter-monomer H-bonds, and water-protein H-bonds. Here, an H-bond is defined by a donor-acceptor (H-O/H_w-O/H-O_w) distance $\leq$ 2.4 Å although many other possible definitions exist. For Gly^{B24}-WI and D-Ala^{B24}, H-bond occupancy in parenthesis corresponds to structures with two different water molecules.

Chain	Donor	Chain	Acceptor	Inter-monomer H-bond occupancy/%	Water-protein H-bond occupancy/%
WT					
monI	Phe^{B24}:H	monII	Tyr^{B26}:O	84	14
monII	Tyr^{B26}:H	monI	Phe^{B24}:O	98	10
monII	Phe^{B24}:H	monI	Tyr^{B26}:O	78	28
monI	Tyr^{B26}:H	monII	Phe^{B24}:O	98	29
GlyB24-SI					
monI	Gly^{B24}:H	monII	Tyr^{B26}:O	81	4
monII	Tyr^{B26}:H	monI	Gly^{B24}:O	96	6
monII	Gly^{B24}:H	monI	Tyr^{B26}:O	95	0
monI	Tyr^{B26}:H	monII	Gly^{B24}:O	95	0
GlyB24-WI					
monI	Gly^{B24}:H	monII	Tyr^{B26}:O	19	30 (28)
monII	Tyr^{B26}:H	monI	Gly^{B24}:O	34	68 (21)
monII	Gly^{B24}:H	monI	Tyr^{B26}:O	0	15 (82)
monI	Tyr^{B26}:H	monII	Gly^{B24}:O	13	56 (32)
AlaB24					
monI	Ala^{B24}:H	monII	Tyr^{B26}:O	92	7
monII	Tyr^{B26}:H	monI	Ala^{B24}:O	99	27
monII	Ala^{B24}:H	monI	Tyr^{B26}:O	38	86
monI	Tyr^{B26}:H	monII	Ala^{B24}:O	87	7
D-AlaB24					
monI	D-Ala^{B24}:H	monII	Tyr^{B26}:O	90	4
monII	Tyr^{B26}:H	monI	D-Ala^{B24}:O	59	46
monII	D-Ala^{B24}:H	monI	Tyr^{B26}:O	0	22 (75)
monI	Tyr^{B26}:H	monII	D-Ala^{B24}:O	73	31 (1)

† The "non-bridging type II" motif (Figure 5b) counts one inter-monomer + one water-protein H-bond. "bridging type I" (Figure 5d) and "bridging type III" (Figure 5f) count two water-protein H-bonds. Because of this the total occupancy can be > 100 %.

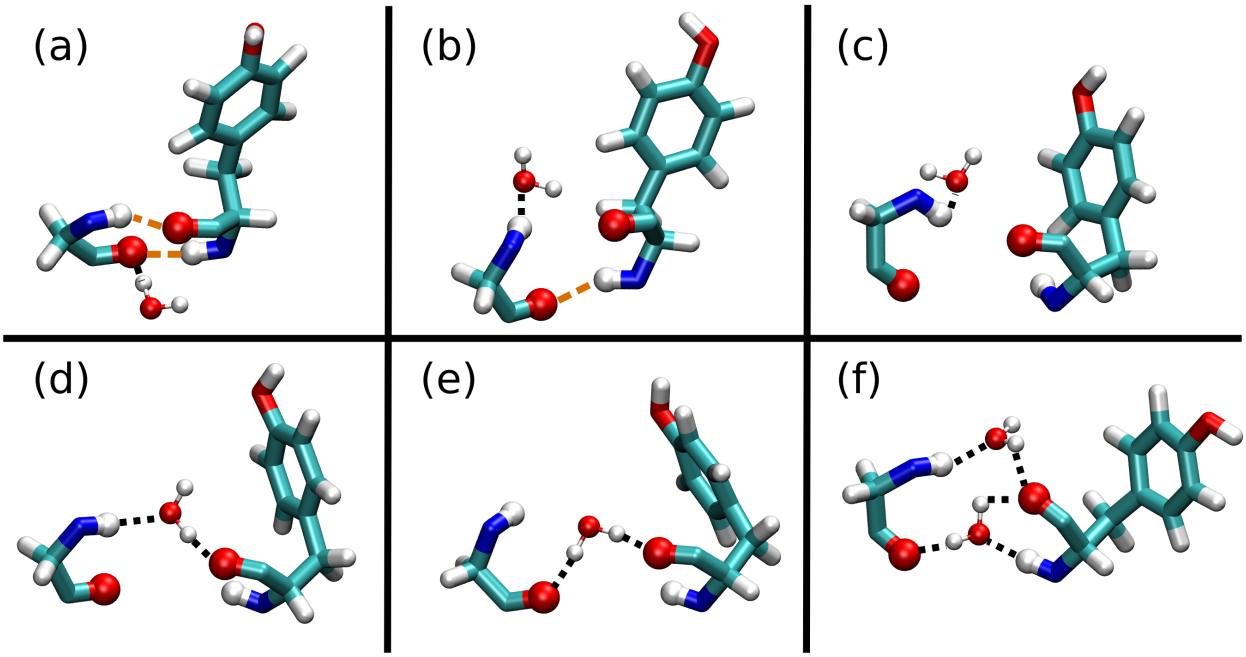

Figure 5: Water molecules interacting at the insulin dimer interface involving B24–B26 inter-monomer H-bond motifs. Only donor and acceptor atoms are highlighted in ball-and-stick representation. H-bonds are depicted by dashed lines (orange for intermonomer, black for water-protein contacts). (a) "non-bridging type I" one water-protein H-bond in the presence of two inter-monomer H-bonds, (b) "non-bridging type II" one water-protein H-bond in the presence of one inter-monomer H-bond, (c) "non-bridging type III" one water-protein H-bond in the absence of inter-monomer H-bonds, (d) and (e) "bridging type I" a single water-bridged H-bond in the absence of inter-monomer H-bonds, (f) "bridging type II" two water-bridged H-bonds in the absence of inter-monomer H-bonds. An H-bond is defined by a donor-acceptor (H-O) distance of 2.4 Å.

respectively (see also Figure S10). For these simulations the H-bonds along the dimerization interface and additional geometrical determinants including the water-occupancy were analyzed. This yielded several types of water-protein H-bonds, shown and described in Figures 5a to f. "Bridging" water-protein H-bonds water have significant interactions with both monomers and thus provide stabilization of the dimer. However, in an MM-GBSA analysis where explicit water molecules do not appear, their influence on the stability is not included. The water-protein H-bond occupancy is higher in Gly^{B24}-WI mutant than in WT and Gly^{B24}-SI (see Table 3). Throughout the MD simulations of Gly^{B24}-WI, consistently one or two water molecules are involved in H-bonds between Gly^{B24} and Tyr^{B26}.

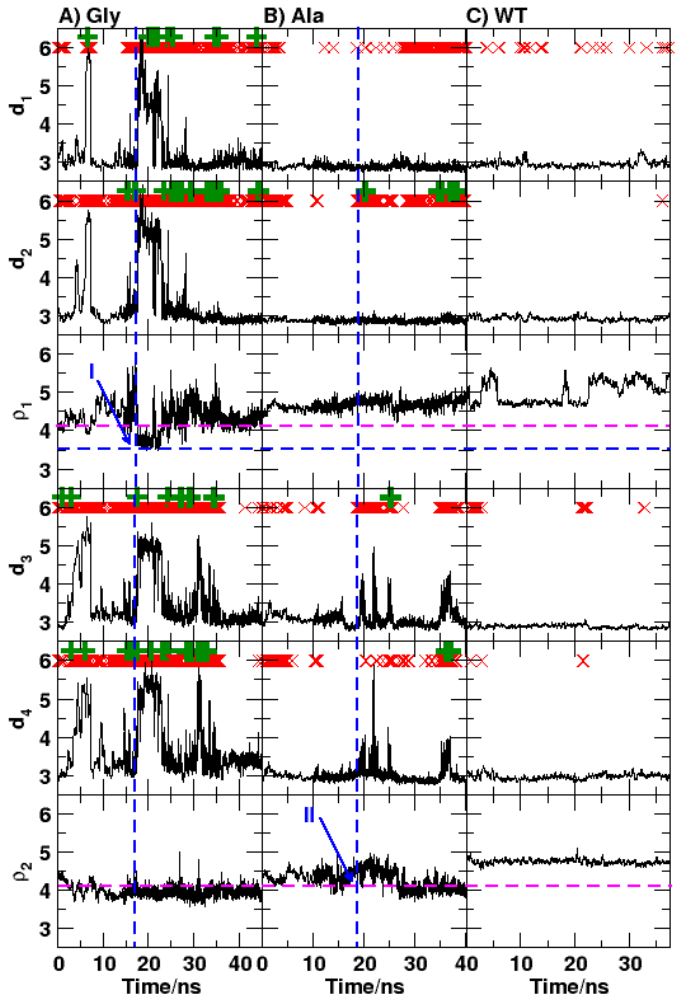

Figure 6: CA-SG (ρ) and intermonomer N-O distances (d, both in Å) as a function of time for A) the glycine mutant, B) the alanine mutant and C) the wildtype. Rows 1, 2, 4 and 5 report the intermonomer hydrogen bonds between residues 24 and $\mathrm{Tyr}^{\mathrm{B26}}$ of the two B-chains. Rows 3 and 6 show the $^{\mathrm{CysB19}}\mathrm{SG}$–$\mathrm{CA}^{\mathrm{XB24}}$ separations for monomers I and II, respectively. Symbols mark times when one (red cross) or two (green plus) water oxygen atoms are within 3.5 Å of both the N- and O-atoms of the corresponding H-bond. Dashed lines are drawn to guide the eye. Generally whenever ρ_1 (or ρ_2) is below 4.1 Å at least one of the intermonomer H-bonds is broken; the magenta dashed line marks this 4.1 Å threshold. The horizontal blue dashed line denotes the minimum distance of ρ_1 that was sampled in the Gly simulation. Vertical dashed lines indicate key points in the simulations. For feature I ρ_1 is at a minimum (0.5 Å below the 4.1 Å threshold) in the Gly simulation and intermonomer bond breakage is clearly observed for all four intermonomer bonds. For feature II - where ρ_1 meets the 4.1 Å threshold for the $\mathrm{Ala}^{\mathrm{B24}}$ mutant and intermonomer bond breakage is observed.

Figure 6 reports several important distance time series to further characterize the dimerization interface. They include inter- and intramonomer distances ρ_1, ρ_2 and d_1 to d_4, see Figure 4B. In Figure 6 the symbols mark times when one (red cross) or two (green plus) water molecules are within 3.5 Å of both atoms of the intermonomer N-O pair. Individual water molecules have also been found to be relevant in simulations of the insulin monomer in water,[65] in HIV-I protease[66,67] or in controlling rebinding of NO to microperoxidase.[68] For Gly^{B24} frequent and spontaneous insertion of one or even two water molecules to replace the direct protein-protein NH–O bond is found. This is sometimes but not exclusively accompanied by losing the NH–O contact and differs considerably for the Ala mutant and the WT protein for which the NH–O Hydrogen bonds are intact for most of the simulation and water molecules are considerably less frequently close to the hydrogen bond. Again, the Gly^{B24} mutant clearly displays a two-state behaviour as already found above: with the NH–O hydrogen bond intact (as is the case for WT and most of the Ala mutant simulations) which corresponds to the SI Gly^{B24} dimer, or with the H-bond broken and typically replaced by a solvent water which is the situation in the WI dimer.

This analysis also provides molecular-level insight into the origins of intermolecular H-bond breaking and water-insertion. The Phe→(Ala,Gly) mutations replace a bulky phenylalanine residue at position 24 by considerably smaller CH_3 (Ala) or H (Gly) moieties (see Figure 7). Hence, the mutations lead to increased conformational freedom of the side chains. This in turn affects the distance(s) ρ_1 (and/or ρ_2), see Figure 6. As a reference, for the WT protein this distance ranges from 4.5 Å to 5.5 Å but decreases to below 4.0 Å for the Gly mutant. For the Ala mutant the separation is closer to the situation in the WT protein. The data in Figure 6 suggests that there are two conditions which lead to breaking of the protein-protein hydrogen bond: (1) the distance $^{CysB19}SG–CA^{X24}$ in at least one monomer, *i.e.* ρ_1 or ρ_2, must be below a threshold of 4.1 Å; and (2) water molecules must be within the vicinity of the intermonomer N-O pair. Furthermore, the data suggests that the smaller ρ_1 or ρ_2, the

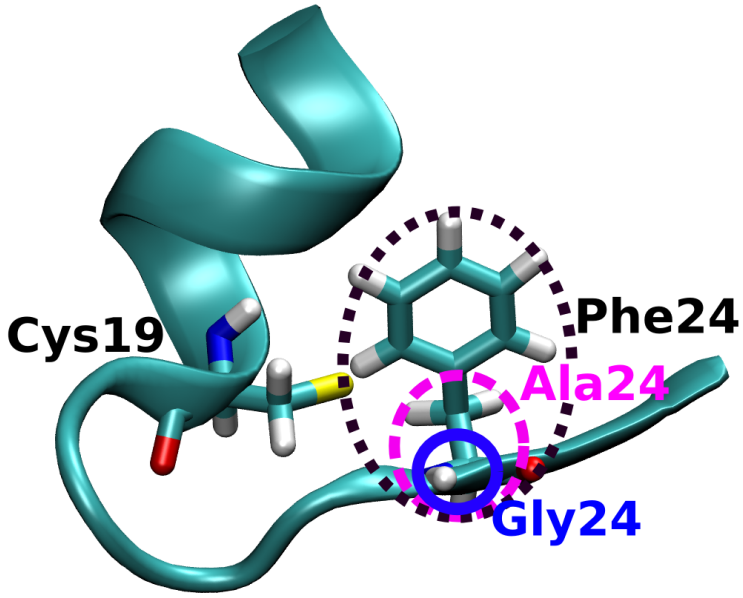

Figure 7: Schematic illustration for the relative sizes of residue 24 in wildtype insulin (Phe^{B24}) and the two mutants (Ala^{B24} and Gly^{B24}). Cys^{B19} is also highlighted to show the variation in steric hindrance between the three cases and to indicate the impact on the key distance, ρ_1 or ρ_2.

larger the distance between the intermonomer N-O pair (*e.g.* 'I' compared to 'II' in Figure 6). Note also that the behavior of the protein-protein H-bonds were closely connected and the breakage of one was often correlated with the breakage of the others.

The present analysis reveals that in Gly^{B24}-WI extensive water-mediated H-bonds entirely replace the β-sheet H-bonds. This is also consistent with the DCCM maps (see Figure S9) where the inter monomer H-bonds are absent for the WI dimer. As a result, the Gly^{B24}-WI dimer has fewer stabilizing inter-monomer interactions resulting in a considerably lower ΔG_{bind} compared to WT and Gly^{B24}-SI within the present MM-GBSA approach. Ala^{B24} has inter-monomer H-bond occupancies similar to WT, whereas D-Ala^{B24} partly differs from WT, in particular one H-bond (D-Ala^{B24}(monomer II):H$\cdots$$Tyr^{B26}$(monomer I):O) is missing throughout the simulation (see Table 3). Consequently, D-Ala^{B24} was found to have a larger number of water-protein H-bonds than Ala^{B24}. The inter-molecular hydrogen bonds, one

of the main factors driving WT insulin dimerization, are absent in the des-Phe^{B25} mutant, thereby preventing its dimerization.

Conclusions

Residue PheB24 plays an essential role in insulin folding, assembly, stability, receptor binding and hormonal signalling.[30] In the present work changes in protein dimer stability resulting from mutations at position B24 were studied using (1) MD simulations with explicit water and (2) free energy simulations using MM-GBSA and TI. To the best of our knowledge, the present work is the first systematic computational study of the relative stabilities and dynamics of dimeric insulin analogues at position B24. The simulations support and extend earlier findings of the importance of residue B24 for the structural integrity of the hormone.[33]

MM-GBSA and TI provide reliable information about the stabilisation of insulin dimers (WT and mutants). Compared to the experimentally determined stabilisation of -7.2 ± 0.8 kcal/mol for the WT, TI finds -8.4 ± 0.2 kcal/mol which is in good quantitative agreement. MM-GBSA is useful for qualitative and comparative purposes but not for quantitative studies. On the other hand, the relative stability changes from TI and MM-GBSA agree quite favourably and suggest that MM-GBSA is useful for ranking the stabilities of WT and mutant dimers.

Substitutions at position B24 of the insulin dimer-forming surface by Gly, Ala and D-Ala amino acid residues give dimeric insulin analogues with reduced dimer stability relative to the WT dimer. The des-PheB25 is exclusively monomeric, as was found by NMR experiments and serves as an additional validation in the present work. The presence of a WI and SI variant of the Gly^{B24} mutant originates from H-bonds which are direct B1$\leftrightarrow$B2 protein-

protein contacts in the SI dimer but water-bridged in the WI dimer and from changes in the orientation of residue B13. This highlights that modifications at one site (B24) can have substantial functional effects. Bridging water molecules replacing H-bonding interactions along dimerization (and oligomerization) interfaces is most likely an ubiquitous feature and should probably be included in estimating dimerization energies which is, however, not routinely done.

Simulations can thus complement existing experiments and provide molecular-level insight for observed differences between chemically related systems. Also, they provide important information for situations in which experiments are technically difficult or impossible as they may be the only direct method to probe the structure, energetics and dynamics at atomic resolution.

Acknowledgement

The authors acknowledge financial support from the University of Basel, and the Swiss National Science Foundation through grant 200020_169079 and the NCCR-MUST.

Supporting Information Available

Binding free energy decomposition data, per residue energetic contribution data and dynamical cross correlation maps. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Baker, E.; Blundell, T.; Cutfield, J.; Cutfield, S.; Dodson, E.; Dodson, G.; Hodgkin, D.; Hubbard, R.; Isaacs, N.; Reynolds, C. The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Phil. Trans. R. Soc. Lond. B Biol.* **1988**, *319*, 369–456.

(2) Jørgensen, A. M. M.; Olsen, H. B.; Balschmidt, P.; Led, J. J. Solution structure of the superactive monomeric des-[Phe (B25)] human insulin mutant: elucidation of the structural basis for the monomerization of des-[Phe (B25)] insulin and the dimerization of native insulin. *J. Mol. Biol.* **1996**, *257*, 684–699.

(3) Brems, D. N.; Alter, L. A.; Beckage, M. J.; Chance, R. E.; DiMarchi, R. D.; Green, L. K.; Long, H. B.; Pekar, A. H.; Shields, J. E.; Frank, B. H. Altering the association properties of insulin by amino acid replacement. *Protein Engineering* **1992**, *5*, 527–533.

(4) Zoete, V.; Meuwly, M.; Karplus, M. Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 79–93.

(5) Pocker, Y.; Biswas, S. B. Self-association of insulin and the role of hydrophobic bonding: a thermodynamic model of insulin dimerization. *Biochemistry* **1981**, *20*, 4354–4361.

(6) El Hage, K.; Pandyarajan, V.; Phillips, N. B.; Smith, B. J.; Menting, J. G.; Whittaker, J.; Lawrence, M. C.; Meuwly, M.; Weiss, M. A. Extending Halogen-based Medicinal Chemistry to Proteins IODO-INSULIN A S A CASE STUDY. *J. Bio. Chem.* **2016**, *291*, 27023+.

(7) Kristensen, C.; Kjeldsen, T.; Wiberg, F. C.; Schäffer, L.; Hach, M.; Havelund, S.; Bass, J.; Steiner, D. F.; Andersen, A. S. Alanine scanning mutagenesis of insulin. *J. Bio. Chem.* **1997**, *272*, 12978–12983.

(8) Hua, Q. X.; Shoelson, S. E.; Kochoyan, M.; Weiss, M. A. Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **1991**, *354*, 238–241.

(9) Derewenda, U.; Derewenda, Z.; Dodson, E.; Dodson, G.; Bing, X.; Markussen, J. X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule: A completely inactive analogue. *J. Mol. Biol.* **1991**, *220*, 425–433.

(10) Vashisth, H.; Abrams, C. F. All-atom structural models of insulin binding to the insulin receptor in the presence of a tandem hormone-binding element. *Proteins: Structure, Function, and Bioinformatics* **2013**, *81*, 1017–1030.

(11) Mirmira, R.; Tager, H. Role of the phenylalanine B24 side chain in directing insulin interaction with its receptor. Importance of main chain conformation. *J. Bio. Chem.* **1989**, *264*, 6349–6354.

(12) Nakagawa, S. H.; Tager, H. S. Importance of aliphatic side-chain structure at positions 2 and 3 of the insulin A chain in insulin-receptor interactions. *Biochemistry* **1992**, *31*, 3204–3214.

(13) Cutfield, J.; Cutfield, S.; Dodson, E.; Dodson, G.; Hodgkin, D.; Reynolds, C. Evidence concerning insulin activity from the structure of a cross-linked derivative. *Hoppe-Seyler´ s Zeitschrift für Physiologische Chemie* **1981**, *362*, 755–762.

(14) Ludvigsen, S.; Olsen, H. B.; Kaarsholm, N. C. A structural switch in a mutant insulin exposes key residues for receptor binding. *J. Mol. Biol.* **1998**, *279*, 1–7.

(15) Nakagawa, S. H.; Tager, H. Role of the phenylalanine B25 side chain in directing insulin interaction with its receptor. Steric and conformational effects. *J. Bio. Chem.* **1986**, *261*, 7332–7341.

(16) Shoelson, S. E.; Lu, Z. X.; Parlautan, L.; Lynch, C. S.; Weiss, M. A. Mutations at the

dimer, hexamer, and receptor-binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions. *Biochemistry* **1992**, *31*, 1757–1767.

(17) Bi, R. C.; Dauter, Z.; Dodson, E.; Dodson, G.; Giordano, F.; Reynolds, C. Insulin's structure as a modified and monomeric molecule. *Biopolymers* **1984**, *23*, 391–395.

(18) Dong, J.; Wan, Z.; Popov, M.; Carey, P. R.; Weiss, M. A. Insulin assembly damps conformational fluctuations: Raman analysis of amide I linewidths in native states and fibrils. *J. Mol. Biol.* **2003**, *330*, 431–442.

(19) Chen, H.; Shi, M.; Guo, Z.-Y.; Tang, Y.-H.; Qiao, Z.-S.; Liang, Z.-H.; Feng, Y.-M. Four new monomeric insulins obtained by alanine scanning the dimer-forming surface of the insulin molecule. *Protein Engineering* **2000**, *13*, 779–782.

(20) Xu, B.; Hu, S.-Q.; Chu, Y.-C.; Huang, K.; Nakagawa, S. H.; Whittaker, J.; Katsoyannis, P. G.; Weiss, M. A. Diabetes-associated mutations in insulin: consecutive residues in the B chain contact distinct domains of the insulin receptor. *Biochemistry* **2004**, *43*, 8356–8372.

(21) Nakagawa, S. H.; Hua, Q.-x.; Hu, S.-Q.; Jia, W.; Wang, S.; Katsoyannis, P. G.; Weiss, M. A. Chiral mutagenesis of insulin contribution of the B20-B23 β-turn to activity and stability. *J. Bio. Chem.* **2006**, *281*, 22386–22396.

(22) Du, H.; Shi, J.; Cui, D.; Zhang, Y. Insulin analogs with B24 or B25 phenylalanine replaced by biphenylalanine. *Acta Biochimica et Biophysica Sinica* **2008**, *40*, 133–139.

(23) Bao, S.-J.; Xie, D.-L.; Zhang, J.-P.; Chang, W.-R.; Liang, D.-C. Crystal structure of desheptapeptide (B24–B30) insulin at 1.6 Å resolution: Implications for receptor binding. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2975–2980.

(24) Shoelson, S.; Fickova, M.; Haneda, M.; Nahum, A.; Musso, G.; Kaiser, E.; Ruben-

stein, A.; Tager, H. Identification of a mutant human insulin predicted to contain a serine-for-phenylalanine substitution. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 7390–7394.

(25) Haneda, M.; Kobayashi, M.; Maegawa, H.; Watanabe, N.; Takata, Y.; Ishibashi, O.; Shigeta, Y.; Inouye, K. Decreased biologic activity and degradation of human [SerB24]-insulin, a second mutant insulin. *Diabetes* **1985**, *34*, 568–573.

(26) Tager, H.; Thomas, N.; Assoian, R.; Rubenstein, A.; Saekow, M.; Olefsky, J.; Kaiser, E. Semisynthesis and biological activity of porcine [LeuB24] insulin and [LeuB25] insulin. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 3181–3185.

(27) Žáková, L.; Kletvíková, E.; Veverka, V.; Lepšík, M.; Watson, C. J.; Turkenburg, J. P.; Jiráček, J.; Brzozowski, A. M. Structural integrity of the B24 site in human insulin is important for hormone functionality. *J. Bio. Chem.* **2013**, *288*, 10230–10240.

(28) Menting, J. G. et al. Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E3395–E3404.

(29) Hua, Q.-X.; Xu, B.; Huang, K.; Hu, S.-Q.; Nakagawa, S.; Jia, W.; Wang, S.; Whittaker, J.; Katsoyannis, P. G.; Weiss, M. A. Enhancing the activity of a protein by stereospecific unfolding conformational life cycle of insulin and its evolutionary origins. *J. Bio. Chem.* **2009**, *284*, 14586–14596.

(30) Pandyarajan, V.; Smith, B. J.; Phillips, N. B.; Whittaker, L.; Cox, G. P.; Wickramasinghe, N.; Menting, J. G.; Wan, Z.-l.; Whittaker, J.; Ismail-Beigi, F.; Lawrence, M. C.; Weiss, M. A. Aromatic anchor at an invariant hormone-receptor interface function of insulin residue b24 with application to protein design. *J. Bio. Chem.* **2014**, *289*, 34709–34727.

(31) Chen, H.; Shi, M.; Guo, Z.-Y.; Tang, Y.-H.; Qiao, Z.-S.; Liang, Z.-H.; Feng, Y.-M. Four new monomeric insulins obtained by alanine scanning the dimer-forming surface of the insulin molecule. *Protein Engineering* **2000**, *13*, 779–782.

(32) DeFelippis, M. R.; Chance, R. E.; Frank, B. H. Insulin self-association and the relationship to pharmacokinetics and pharmacodynamics. *Critical Reviews in Therapeutic Drug Carrier Systems* **2001**, *18*.

(33) Antolíková, E.; Žáková, L.; Turkenburg, J. P.; Watson, C. J.; Hančlová, I.; Šanda, M.; Cooper, A.; Kraus, T.; Brzozowski, A. M.; Jiráček, J. Non-equivalent role of inter-and intramolecular hydrogen bonds in the insulin dimer interface. *J. Bio. Chem.* **2011**, *286*, 36968–36977.

(34) Jiracek, J.; Zakova, L.; Antolikova, E.; Watson, C. J.; Turkenburg, J. P.; Dodson, G. G.; Brzozowski, A. M. Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1966–1970.

(35) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673–7697.

(36) Zoete, V.; Meuwly, M.; Karplus, M. A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer. *J. Mol. Biol.* **2004**, *342*, 913–929.

(37) Zoete, V.; Meuwly, M. Importance of individual side chains for the stability of a protein fold: computational alanine scanning of the insulin monomer. *J. Comput. Chem.* **2006**, *27*, 1843–1857.

(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(39) MacKerell, Jr., A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(40) Mackerell, A.; Feig, M.; Brooks, C. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(41) MacKerell, A.; Feig, M.; Brooks, C. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

(42) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1âĂŞ2*, 19 – 25.

(43) Klauda, J. B.; Venable, R. M.; Freites, J. A.; OâĂŹConnor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

(44) Bernstein, F.; Koetzle, T.; Williams, G. J.; Jr., E. F. M.; Brice, M.; Rodgers, J.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.

(45) Berman, H. M. et al. The protein data bank. *Biological Crystallography* **2002**, *58*, 899–907.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(47) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(48) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston me thod. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(49) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(50) Kollman, P.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.; Cheatham, T. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(51) Gohlke, H.; Kiel, C.; Case, D. A. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *J. Mol. Biol.* **2003**, *330*, 891–913.

(52) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.

(53) Boresch, S. The role of bonded energy terms in free energy simulations-Insights from analytical results. *Mol. Sim.* **2002**, *28*, 13–37.

(54) Boresch, S.; Karplus, M. The role of bonded terms in free energy simulations: 1. Theoretical analysis. *J. Phys. Chem. A* **1999**, *103*, 103–118.

(55) Boresch, S.; Karplus, M. The Role of Bonded Terms in Free Energy Simulations. 2. Calculation of Their Influence on Free Energy Differences of Solvation. *J. Phys. Chem. A* **1999**, *103*, 119–136.

(56) Shobana,; Roux, B.; Andersen, O. S. Free Energy Simulations:âĂĽ Thermodynamic Reversibility and Variability. *J. Phys. Chem. B* **2000**, *104*, 5179–5190.

(57) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

(58) Bash, P.; Singh, U.; Brown, F.; Langridge, R.; Kollman, P. Calculation of the relative change in binding free energy of a protein-inhibitor complex. **1987**, *235*, 574–576.

(59) Pearlman, D. A Comparison of Alternative Approaches to Free-Energy Calculations. *J. Phys. Chem.* **1994**, *98*, 1487–1493.

(60) Strazza, S.; Hunter, R.; Walker, E.; Darnall, D. W. The thermodynamics of bovine and porcine insulin and proinsulin association determined by concentration difference spectroscopy. *Archives of Biochemistry and Biophysics* **1985**, *238*, 30–42.

(61) Tidor, B.; Karplus, M. The contribution of vibrational entropy to molecular association: the dimerization of insulin. *J. Mol. Biol.* **1994**, *238*, 405–414.

(62) Yao, Z.-P.; Zeng, Z.-H.; Li, H.-M.; Zhang, Y.; Feng, Y.-M.; Wang, D.-C. Structure of an insulin dimer in an orthorhombic crystal: the structure analysis of a human insulin mutant (B9 SerGlu). *Acta Crystallographica Section D: Biological Crystallography* **1999**, *55*, 1524–1532.

(63) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. J. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.* **2011**, *83*, 1637–1641.

(64) Domagala, M.; Grabowski, S. C-(HN)-N-... and C-(HS)-S-... hydrogen bonds - Influence of hybridization on their strength. *J. Phys. Chem. A* **2005**, *109*, 5683–5688.

(65) Bagchi, K.; Roy, S. Sensitivity of Water Dynamics to Biologically Significant Surfaces of Monomeric Insulin: Role of Topology and Electrostatic Interactions. *J. Phys. Chem. B* **2014**, *118*, 3805–3813.

(66) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D.; Kellogg, G. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J. Medic. Chem.* **2004**, *47*, 4507–4516.

(67) Thorsteinsdottir, H. B.; Schwede, T.; Zoete, V.; Meuwly, M. How inaccuracies in protein structure models affect estimates of protein-ligand interactions: Computational analysis of HIV-1 protease inhibitor binding. **2006**, *65*, 407–423.

(68) Mondal, P.; Meuwly, M. Solvent Composition Drives the Rebinding Kinetics of Nitric Oxide to Microperoxidase. *Sci. Rep.* **2018**, *8*.

Graphical TOC Entry