

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Probing the Differential Dynamics of Monomeric and Dimeric Insulin from Amide-I IR Spectroscopy

Journal:	<i>The Journal of Physical Chemistry</i>
Manuscript ID	jp-2019-046288.R1
Manuscript Type:	Article
Date Submitted by the Author:	22-Jun-2019
Complete List of Authors:	Desmond, Jasmine; Universitat Basel Departement Chemie, Koner, Debasish; University of Basel, Chemistry Meuwly, Markus; Universitat Basel, Chemistry

SCHOLARONE™
Manuscripts

Probing the Differential Dynamics of Monomeric and Dimeric Insulin from Amide-I IR Spectroscopy

Jasmine L. Desmond, Debasish Koner, and Markus Meuwly*

*Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel,
Switzerland*

E-mail: m.meuwly@unibas.ch

Abstract

The monomer \leftrightarrow dimer equilibrium for insulin is one of the essential steps in forming the receptor binding-competent monomeric form of the hormone. Despite this importance, the thermodynamic stability - in particular for modified insulins - is quite poorly understood in part due to experimental difficulties. This work explores 1- and 2-dimensional infrared spectroscopy in the range of the amide-I band for the hydrated monomeric and dimeric wild type hormone. It is found that for the monomer the frequency fluctuation correlation function (FFCF) and the 1d-infrared spectra are position sensitive. The spectra for the -CO probes at the dimerization interface (residues Phe24, Phe25, Tyr26) split and indicate an asymmetry despite the overall (formal) point symmetry of the dimer structure. Also, the decay times of the FFCF for the same -CO probe in the monomer and the dimer can differ by up to one order of magnitude, for example for residue PheB24 which is solvent exposed for the monomer but at the interface for the dimer. The spectroscopic shifts correlate approximately with the average number of hydration waters and the magnitude of the FFCF at time zero. However,

1
2
3 this correlation is only qualitative due to the heterogeneous and highly dynamical en-
4 vironment. Based on density functional theory calculations the dominant contribution
5 for solvent-exposed -CO is found to arise from the surrounding water ($\sim 75\%$) whereas
6 the protein environment contributes considerably less. The results suggest that infrared
7 spectroscopy is a positionally sensitive probe of insulin dimerization, in particular in
8 conjunction with isotopic labeling of the probe.
9
10
11
12
13
14
15
16

17 1 Introduction

18
19
20 Insulin is a small, aggregating protein with an essential role in regulating glucose uptake
21 in cells. The WT hormone crystallizes as a hexamer consisting of three dimers with either
22 two or four Zn atoms bound to it.¹ The dimer consists of two monomers (chains A and
23 B with 21 and 30 amino acids, respectively), connected by two inter-chain (Cys^{A7}-Cys^{B7},
24 Cys^{A20}-Cys^{B19}) and one intra-chain (Cys^{A6}-Cys^{A11}) disulfide bonds. Under physiological
25 conditions the thermodynamically favoured state is the insulin dimer primarily stabilized
26 by hydrogen bonds between residues B24-B26 of one monomer and B26-B24 of the sec-
27 ond monomer.¹⁻⁵ However, the functionally relevant state is the monomer which can bind to
28 the insulin receptor. Therefore, dissociation of the dimer is of great physiological importance.
29
30
31
32
33
34
35
36
37
38
39

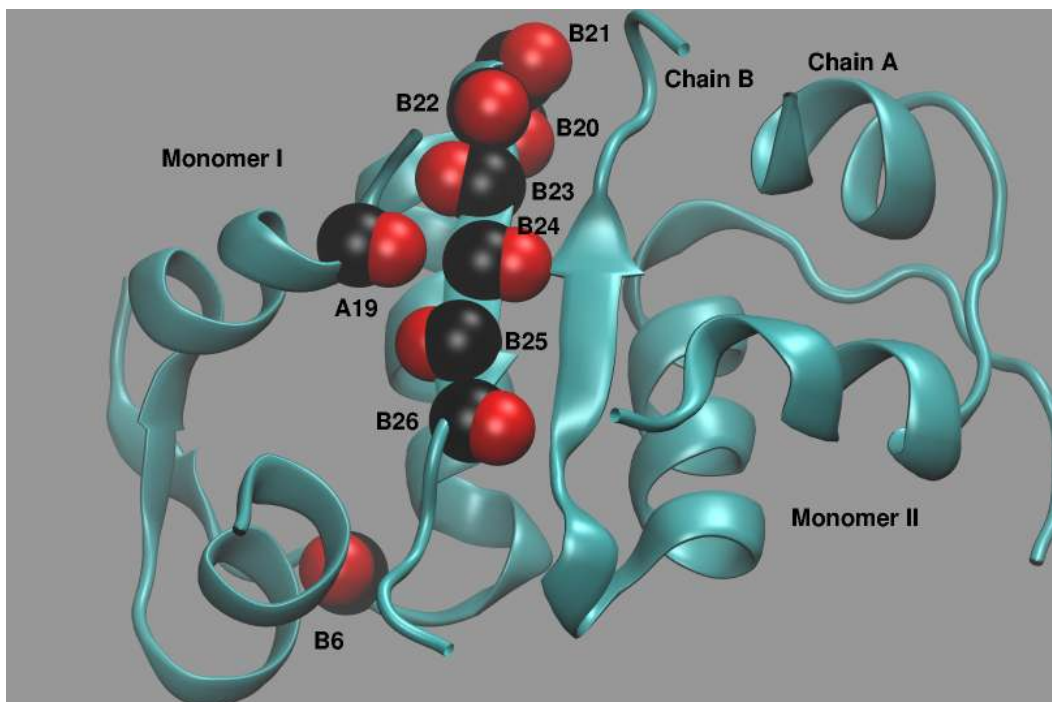
40 However, experimental data for dimerization free energies of insulin analogues is scarce due
41 to several challenges. The measured dimerization energy for WT insulin is -7.2 kcal/mol in
42 favour of the dimer.⁶ Apart from this, it was found that the Gly^{B24} mutant does not dimerize
43 in aqueous solution at pH 1.9.⁷ Alanine scanning of the dimerization interface showed that
44 the Ala^{B24} analogue is monomeric and does not readily aggregate.^{8,9} Furthermore, isother-
45 mal titration calorimetry (ITC) measurements of N-methylated insulin dimer analogues at
46 positions B24, B25, and B26^{10,11} revealed considerably reduced (by a factor of 5) dimeriza-
47 tion capabilities compared with human insulin. Also, using computations the dimerization
48 free energies for the WT dimer^{4,12} and, more recently, for several mutants at position B24
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 were determined.¹³
4
5
6

7 The current view of the insulin structure-function relationship is derived primarily from
8 insulin hexamer and dimer crystal structures, as well as from studies of the structure-
9 activity correlations of chemically modified and/or naturally occurring mutant insulins in
10 solution.^{7,14-21} Mutagenesis of the dimer-forming surface of insulin can yield analogues with
11 a reduced tendency to aggregate and pronounced differences in the pharmacokinetic proper-
12 ties with potentially promising therapeutic applications.^{14,22} Typical experimental methods
13 for quantitative studies of insulin dimerization are ITC³ or NMR spectroscopy. ITC requires
14 relatively high protein concentrations, while NMR spectroscopy can be (too) slow for such
15 purposes. Based on transition state theory a free energy difference of -7.2 kcal/mol between
16 the WT dimer and two separate monomers corresponds to a sub- μ s time scale. Several NMR
17 studies of active monomeric insulin mutants show a rearrangement of the C-terminal end of
18 chain B.^{7,20} A combined Raman spectroscopy and microscopy study of insulin in different
19 aggregation states (monomer, dimer, hexamer and fibril) shows that dimerization damps
20 fluctuations at an intermolecular β -sheet.²³ More recently, vibrational spectroscopy has also
21 been used to characterize the insulin dimer²⁴ and isotopically labelled insulins.²⁵
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 Earlier experiments on model peptides in solution (water), such as the blocked alanine trimer
40 (Ala)₃ already found that the two amide-I bands appear at positions 25 cm⁻¹ apart.²⁶ Similar
41 observations have been made for acyl-proline-NH₂ in chloroform for which the two oscillators
42 absorb at maximum frequencies 40 cm⁻¹ to 50 cm⁻¹ apart depending on the time delay in
43 the 2-dimensional infrared experiments.²⁷ Therefore, it is expected that by increasing the
44 number of reporters, e.g. in a protein with one amide-I probe per amino acid, the -CO
45 absorption bands will partly overlap and cover an even larger range due to differing chemical
46 environments that are probed during the dynamics.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Given the fundamental importance of insulin dimerization it is desirable to develop method-
4 ologies that are sensitive to and provide direct information on the aggregation state of the
5 hormone in solution. In the present work vibrational spectroscopy is used for that. Infrared
6 spectroscopy (IR) has been shown to be environment specific. For example, the CH-stretch
7 in HCO^+ is known to shift depending on the number of surrounding argon atoms.²⁸ In
8 Ar-HCO^+ the red shift is -274 cm^{-1} whereas for each additional Ar atom an incremental
9 blue shift by $\sim 10 \text{ cm}^{-1}$ is observed. For the first complete solvation sphere, containing 12
10 Argon atoms, the red shift relative to the gas phase reduces to -155 cm^{-1} . Hence vibrational
11 spectroscopy is a sensitive environmental probe of a spectroscopic reporter.
12
13
14
15
16
17
18
19
20
21
22



23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46 Figure 1: Structure of insulin dimer (PDB Code: 4INS) consisting of monomer I and II, each
47 with chains A and B. Some -CO probes are explicitly represented as van der Waals spheres
48 and labelled.
49

50
51 Because the monomer \leftrightarrow dimer equilibrium is one of the essential steps in forming the re-
52 ceptor binding-competent monomeric form of insulin and the process is difficult to study
53 quantitatively by experiments, MD simulations are an attractive alternative to characterize
54
55
56
57
58
59
60

1
2
3 the stability of insulin dimers. In the present work the amide-I vibration for monomeric
4 and dimeric WT insulin is studied using validated energy functions for the -CO probe with
5 the aim to quantify the differences in the spectroscopy and dynamics of the two aggregation
6 states. However, direct comparison of absolute frequencies for insulin in solution is difficult
7 because typical spectroscopic investigations are carried out in solutions with up to 20 %
8 ethanol, HCl and NaCl and a pronounced sensitivity of the observables depending on the
9 amount of cosolvent has been reported.²⁴ Under such circumstances it appears to be more
10 meaningful to report frequency shifts relative to a reference frequency. This is different, e.g.,
11 for (Ala)₃ for which experiments in pure water have been carried out²⁹ and direct comparison
12 is possible (see below).
13
14
15
16
17
18
19
20
21
22
23
24

25 The frequency trajectory of a local reporter can be followed in different ways. One of them
26 uses so-called parametrized “frequency maps”, pioneered by Skinner and coworkers, which
27 are precomputed for a given reporter from a large number of ab initio calculations.^{30,31} Such
28 maps are very convenient and, once parametrized, computationally advantageous. Much
29 has been learned from them on the multidimensional spectroscopy of water and for peptides
30 and proteins in the range of the amide-I band.^{30,32} To develop such a map, snapshots from
31 MD simulations are used. The frequencies of the reporter are computed using electronic
32 structure methods (e.g. B3LYP/6-311++G**) by stretching and compressing the bond
33 of interest to map out a 1-dimensional potential energy curve from which the 1-0 and 2-1
34 transitions are calculated by solving the 1-dimensional Schrödinger equation. At the same
35 time, the strength of the electric field at the water hydrogen atom (for water-OH maps)³¹ is
36 determined and the frequency and field strength are treated as a set. The ensuing, empiri-
37 cal correlation is then fit to a parametric function which allows to determine the expected
38 frequency of the parametrized oscillator from the knowledge of the strength of the electric
39 field alone.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 While convenient, such a treatment also introduces approximations. The frequency shift for
4 given field strength has typically a finite width which can be up to 50 cm^{-1} for OH.³¹ Hence,
5 such a map can be regarded as a “mean field approximation” as the fluctuations are usually
6 disregarded. Often, the field is assumed to be linear along the probe, although exceptions
7 exist.³³ As the electric fields on molecular length scales can be strong, corrections due to the
8 interaction of the charges with the field gradient and higher order terms may be required for
9 an improved relationship between oscillator frequency and field strength. The clusters used
10 in the electronic structure calculations are often quite small (e.g. 4 \AA within the probe with
11 implicit inclusion of additional waters up to 8 \AA)³¹ out of necessity to keep the electronic
12 structure calculations at a manageable level. In practice, the maps are fitted to reproduce
13 experimentally determined line positions but the final models are transferable to some de-
14 gree.³² A potential matter of concern of such maps is the fact that the nuclear dynamics is
15 followed with an energy function that is typically that of an empirical force field whereas the
16 analysis of the snapshots is carried out at a different level of theory. Thus, there is a con-
17 ceptual difficulty in that the distribution of conformations from the MD simulations is not
18 necessarily consistent with that one would obtain if sampling would have been carried out
19 at the DFT or even MP2 level of theory which is the level at which the maps are determined.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 Alternatively, the sampling of the configurations and computing frequencies for given snap-
40 shots can also be done by using the same energy function (“scan”). This has been pursued
41 in the past in the context of multipolar force fields and applied to systems such as hydrated
42 CN^- ,³⁴ the -CO probe in N-methyl acetamide,³⁵ or N_3^- in the gas phase and in solution.³⁶
43 In this approach the MD simulations are carried out with the same energy function that
44 is also used for the analysis, which is typically a multipolar or a fluctuating charge repre-
45 sentation for the electrostatics around the spectroscopic probe and an anharmonic (Morse)
46 or even superior reproducing kernel PES³⁶⁻³⁸ for the bonded terms (stretch and bend). On
47 each snapshot the local frequency is determined from either an instantaneous normal mode
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 calculation or from solving the 1d or 3d nuclear Schrödinger equation.^{34,36} For the latter
4 approach, the PES is scanned along all relevant degrees of freedom, similar to the map ap-
5 proach. However, when solving the Schrödinger equation, choosing the reduced mass is not
6 obvious, in particular for covalently bound probes.
7
8
9

10
11
12 For CN^- in solution, decay times of the FFCF and the full width at half maximum are in
13 quantitative agreement with experiment using such an approach.³⁴ Similarly, for NMA in so-
14 lution the decay times of the FFCF have been found to agree well with experiment, whereas
15 the full width at half maximum is underestimated (14 cm^{-1} compared with 29 cm^{-1}).³⁵ As a
16 comparison, a more recent study of the dynamics of NMAD in D_2O based on a parametrized
17 DFT map finds a FWHM of 43 cm^{-1} which overestimates the experimentally determined
18 value.³⁹ Conversely, a more elaborate procedure which fits³² the map to experimental data
19 in different solvents yields a FWHM of 34 cm^{-1} .
20
21
22
23
24
25
26
27
28
29
30

31 Overall, both approaches (“map” and “scan”) make their own approximations and neither
32 of them is strictly “exact”. In the following, we use instantaneous normal modes and the
33 explicit solution of the nuclear Schrödinger equation to follow the frequency trajectory of
34 the oscillator in the field of its environment.
35
36
37
38
39
40

41 This article is organized as follows: Section 2 introduces the methods. In Section 3 first re-
42 sults on the spectroscopy of hydrated N-Methyl-Acetamide (NMA) in solution are reported.
43 This is followed by the discussion of the frequency fluctuation correlation functions and line-
44 shapes of $-\text{CO}$ in monomeric and dimeric insulin in solution. Finally, conclusions are drawn
45 in Section 4.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2 Methods

2.1 Molecular Dynamics Simulations

All molecular dynamics (MD) simulations were carried out using the CHARMM MD simulation package,⁴⁰ with the CHARMM22 force-field.⁴¹ The X-ray crystal structure of the insulin dimer (PDB code 4INS)¹ was solvated in a cubic box ($52.77 \times 52.77 \times 52.77 \text{ \AA}^3$) of TIP3P⁴² water molecules, which leads to a total system size of ~ 17500 atoms. For the monomer simulations chains A and B were retained and also solvated in a water box ($52.77 \times 52.77 \times 52.77 \text{ \AA}^3$, the same box size as the dimer). Hydrogen atoms were included using CHARMM. The systems were neutralised with potassium ions (one for the monomeric system and two for the dimer).

The systems were minimised, heated to the relevant temperature for 50 ps and equilibrated for 250 ps in the *NVT* ensemble. Production data (up to 4 ns) was generated in the *NPT* ensemble, with coordinates saved every 5 fs for subsequent analysis. All bonds involving hydrogen atoms were constrained using SHAKE.⁴³ A Velocity Verlet integrator⁴⁴ and Nose-Hoover thermostat^{45,46} were employed in the *NVT* simulations. For the *NPT* simulations an Andersen and Nose-Hoover constant pressure and temperature algorithm was used⁴⁶⁻⁴⁸ together with a leapfrog integrator.⁴⁹ Non-bonded interactions were treated with a switching function between 10 \AA and 14 \AA .⁵⁰ The Particle Mesh Ewald (PME) method was used for the electrostatic interactions.⁵¹

Two simulations of N-methyl acetamide (NMA) - featuring an amide bond with (1) natural $^{12}\text{C}^{16}\text{O}$ isotopes and (2) heavy $^{13}\text{C}^{18}\text{O}$ isotopes - were carried out to validate the computational procedures. In these simulations the entire NMA molecule was treated with multipoles.^{35,52-54} For the bonded terms the parameters were those of the CHARMM22 force field except for the -CO stretch for which a Morse potential $V(r) = D_e(1 - \exp(-\beta(r - r_e)))^2$ was

used. The parameters are $D_e = 120.47$ kcal/mol, $\beta = 2.174$ Å⁻¹ and $r_0 = 1.294$ Å which reproduce the experimental gas phase C=O stretch frequency of 1731 cm⁻¹.^{35,55}

2.2 Calculation of the 1D infrared spectra

The anharmonic frequencies were computed from the time-ordered snapshots extracted from the MD simulation every 5 fs. For this purpose, 1-dimensional potential energy curves were calculated along the -CO bond for each snapshot by freezing all other atoms of the system. The anharmonic transition frequencies ($v = 0 \rightarrow 1$) were then calculated by solving the 1D time-independent Schrödinger equation using the LEVEL⁵⁶ program to obtain the frequency trajectory. Following this, the frequency fluctuation correlation function, $C(t) = \langle \delta\omega(0)\delta\omega(t) \rangle$, was calculated. Here, $\delta\omega(t) = \omega(t) - \langle \omega(t) \rangle$, where $\langle \omega(t) \rangle$ is the ensemble average. The lineshape function $g(t)$ is determined from the FFCE within the cumulant approximation as,⁵⁷

$$g(t) = \int_0^t \int_0^{\tau'} \langle \delta\omega(\tau'')\delta\omega(0) \rangle d\tau'' d\tau'. \quad (1)$$

Depending on how $C(t)$ is represented, the integral in Eq. 1 can be evaluated analytically. Two parametrizations were considered for this, depending on the shape of $C(t)$: either a superposition of exponentially decaying functions⁵⁸

$$C(t) = \sum_{i=1}^2 a_i e^{-t/\tau_i} + \Delta_0, \quad (2)$$

or one including a recurrence term⁵⁹

$$C(t) = a_1 \cos(\gamma t) e^{-t/\tau_1} + \sum_{i=2}^3 a_i e^{-t/\tau_i} + \Delta_0 \quad (3)$$

where the a_i are amplitudes, the τ_i the decay constants and Δ_0 a constant for cases in which $C(t)$ does not decay to zero for the correlation times considered. The decay times τ_i of the

1
2
3 frequency fluctuation correlation function reflect the characteristic time-scales of the solvent
4 fluctuations to which the solute degrees of freedom are coupled.
5
6
7

8
9 Finally, the 1D IR spectra is calculated as
10

$$11 \quad I(\omega) = 2\Re \int_0^\infty e^{i(\omega - \langle \omega \rangle)t} e^{-g(t)} e^{-\frac{t\alpha}{2T_1}}, \quad (4)$$

12
13
14
15

16 where $\langle \omega \rangle$ is the average transition frequency obtained from the distribution, $T_1 = 0.45$ ps
17 is the vibrational relaxation time and $\alpha = 0.5$ is a phenomenological factor to account for
18 lifetime broadening.⁵⁷
19
20
21
22

23
24 The current analysis makes a few assumptions in calculating the FFCF and the 1D IR spec-
25 tra. Using the cumulant approximation was thoroughly tested within the present framework
26 for CN^- in solution for which the lineshapes from analysis with and without the cumulant
27 approximation differ by less than 10 %. Since CN^- in solution is more strongly affected
28 by coupling between rotation and vibration (approximations in the orientational dynamics),
29 this effect is expected to be smaller in the present case. Furthermore, previous work has
30 found that non-Condon effects for amide-I vibrations are small.^{39,60}
31
32
33
34
35
36
37
38
39
40

41 **3 Results and Discussion**

42
43

44 **3.1 NMA (N-Methylacetamide) in Water**

45
46

47 In order to validate the present computational approach, simulations of N-Methylacetamide
48 (NMA) in water for the natural ($\text{C}^{12}\text{O}^{16}$) and the heavy ($^{13}\text{C}^{18}\text{O}$) isotopes in the amide
49 bond were carried out. NMA is a suitable model for the peptide linkages of the protein
50 backbone and therefore a meaningful proxy. Previous experimental measurements of the
51 CO absorption band of NMA in the gas phase found absorptions in the range of 1723 cm^{-1}
52
53
54
55
56
57
58
59
60

1
2
3 to 1731 cm^{-1} (average of $\sim 1725\text{ cm}^{-1}$)^{55,61,62} which shifts to the red (1617 cm^{-1} to 1625
4 cm^{-1})⁶²⁻⁶⁴ for NMA in solution. Hence, the experimentally observed red shift between gas-
5 and condensed-phase is $\sim 100\text{ cm}^{-1}$. Previous simulations in which NMA was kept frozen in
6 the simulations and the frequency trajectories were determined from scanning the potential
7 energy for each snapshot along the amide-I normal mode found a red shift of 19 cm^{-1} .³⁵
8
9
10
11
12
13
14

15 In the present work the chromophores are fully flexible and frequencies were determined
16 either from an instantaneous normal mode (INM) analysis (for NMA) or from solving the
17 1-dimensional Schrödinger equation (for NMA and the -CO probes in insulin monomer and
18 dimer) for the potential energy scanned along the -CO local mode. In NMA, there are two
19 normal modes with large contributions of the -CO stretch; one at 1592 cm^{-1} (see Figure
20 2B) and the second one at 1801 cm^{-1} (Figure S1). Although the second mode is closer in
21 frequency to the experimentally observed band,⁵⁵ the first mode was followed as its -CO local
22 mode character was larger when analyzing the normal modes of NMA in solution. However,
23 as only *relative* frequency shifts depending on the environmental dynamics are of interest
24 in the present work the results are not expected to depend on this choice when using the
25 instantaneous normal mode analysis.
26
27
28
29
30
31
32
33
34
35
36
37
38

39 The power spectrum for NMA in the gas phase is reported in Figure 2A. For the natural
40 isotope (green) the maximum peak frequency is at 1724 cm^{-1} , in good agreement with ex-
41 periment, which shifts to 1698 cm^{-1} for simulations in water (black) and corresponds to a
42 solvent-induced red shift of -26 cm^{-1} . The isotope shift (red) for simulations in solution is
43 65 cm^{-1} , compared with 60 cm^{-1} from experiment.⁶⁵ The absolute red shift in going from
44 the gas phase into solution for the amide-I vibration between computations ($\sim 30\text{ cm}^{-1}$)
45 and experiment ($\sim 100\text{ cm}^{-1}$) differs by about 70 cm^{-1} . This can be explained by the
46 fact that in the present simulations the -CO stretch potential for simulations in the gas- and
47 condensed-phase is identical. Thus, the computed frequency shift is entirely due to the differ-
48
49
50
51
52
53
54
55
56
57
58
59
60

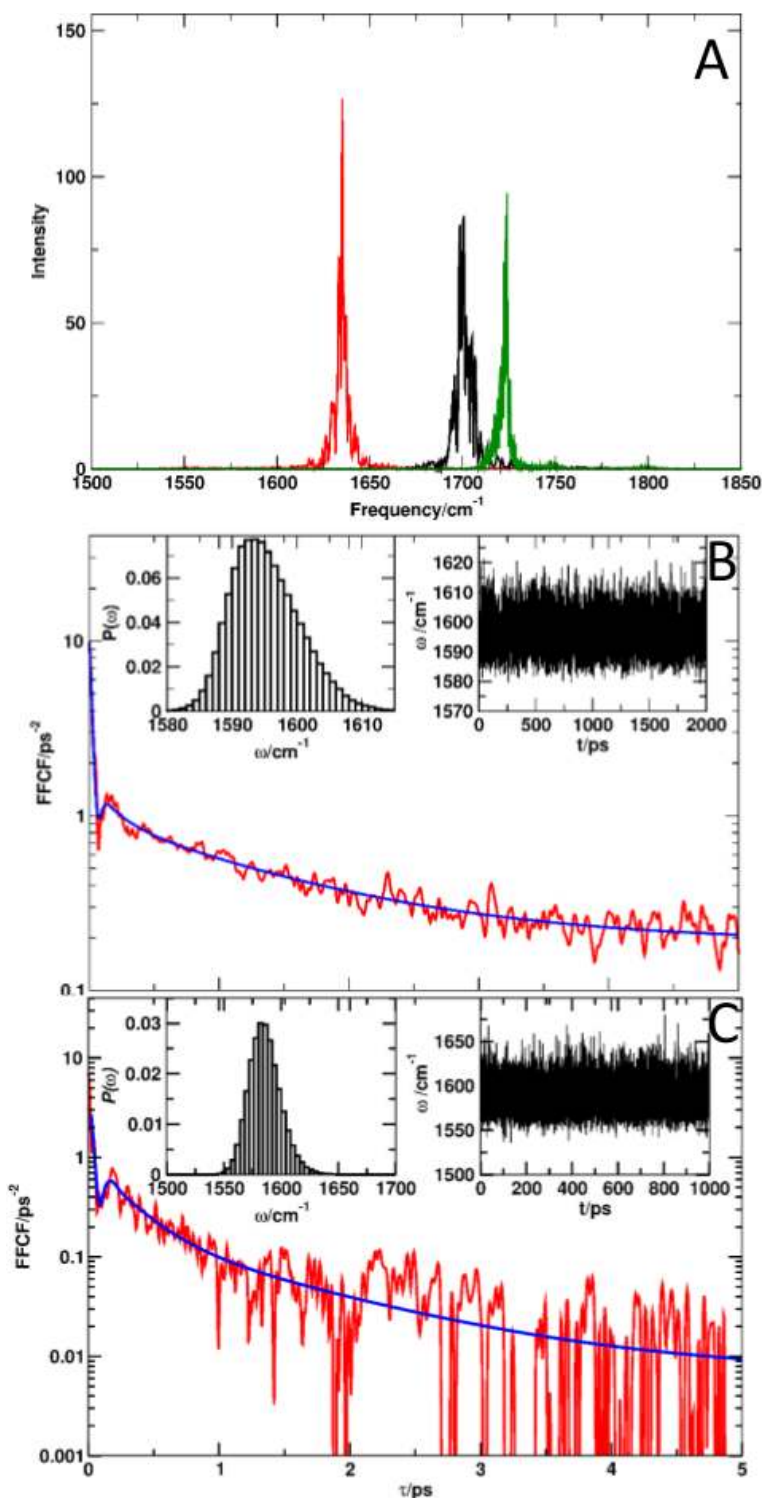


Figure 2: Power spectra and frequency fluctuation correlation functions (FFCFs) for NMA. Panel A: power spectra for ¹²C¹⁶O (green, in gas phase, $\nu_{\text{max}} = 1723 \text{ cm}^{-1}$), ¹²C¹⁶O (black, in solution, $\nu_{\text{max}} = 1700 \text{ cm}^{-1}$) and ¹³C¹⁸O (red, $\nu_{\text{max}} = 1635 \text{ cm}^{-1}$). Panel B: FFCF from INM analysis for ¹²C¹⁶O. The insets report the frequency distribution and frequency trajectory $\omega(t)$. Panel C: FFCF with frequencies calculated using LEVEL for ¹²C¹⁶O together with the frequency distribution and the frequency trajectory. The parameters used for the fits are given in Table 1. In panels B and C the red lines are the raw data and the blue lines are the fit.

ent electrostatic interactions in the two environments. It is, however, expected that the -CO stretch potentials in the gas- and condensed phase slightly differ which could, in principle, be captured by the parametrization. As an illustration, changing the repulsive parameter β in the Morse potential from 2.174 \AA^{-1} by $\sim 0.12 \text{ \AA}^{-1}$ to 2.050 \AA^{-1} shifts the maximum of the power spectrum in solution by -70 cm^{-1} , to the value known from experiment. Likewise, changing the van der Waals radii of the carbon and oxygen atoms by 5 % shifts the absorption frequency by $\pm 15 \text{ cm}^{-1}$. Nevertheless, because in the present work relative frequency shifts in the same environment (water) are of interest this was not pursued further.

Computing frequencies from INM or from scanning the CO-potential in the solvent environment and solving the 1-dimensional Schrödinger equation provides alternative ways to construct the frequency trajectory $\omega(t)$ from which the frequency fluctuation correlation function (FFCF) is determined. Both methods make somewhat different approximations. When using instantaneous normal modes, the harmonic approximation is assumed and the motions are not necessarily local, i.e. a particular normal mode will contain CO stretch motion coupled to other internal degrees of freedom. On the other hand, solving for the stationary states by scanning the CO-potential assumes a local mode and requires one to choose a reduced mass which, however, only affects the absolute frequencies but leaves relative blue- and red-shifts unchanged. To further test this, the frequency distribution was analyzed from solving the 1-dimensional Schrödinger equation for 10^5 snapshots by scanning the potential along the amide-I normal mode. The frequency distribution (see Figure S2) peaks at 1645 cm^{-1} which is shifted by 20 cm^{-1} to the blue compared with experiment at 1625 cm^{-1} .

The maximum of the frequency distribution $P(\omega)$ calculated using INM for NMA in water (Figure 2B) is at 1594 cm^{-1} . For the frequency trajectory from solving the 1D Schrödinger equation (Figure 2C) the maximum is found at 1586 cm^{-1} . In these calculations the reduced

Table 1: Fitting parameters from exponential fits of the FFCFs for $^{12}\text{C}^{16}\text{O}$ and $^{13}\text{C}^{18}\text{O}$ in NMA. Amplitudes a_i and offset Δ_0 are reported in cm^{-2} , τ_i are in ps and γ in ps^{-1} . The short time scale from the experiments in Ref.²⁹ (50 to 100 fs) was estimated. In Ref.³⁵ only a sum of three exponentials was used to fit the FFCF.

	a_1 ps ⁻²	τ_1 ps	γ ps ⁻¹	a_2 ps ⁻²	τ_2 ps	a_3 ps ⁻²	τ_3 ps	Δ_0 ps ⁻²
INM ³⁵	3.89	0.07	–	1.04	0.42	1.01	1.07	–
mode 33 (INM)	8.38	0.03	32.41	0.58	0.18	0.79	1.34	0.19
$^{12}\text{C}^{16}\text{O}$ (1d-SE)	3.79	0.04	29.38	0.77	0.24	0.22	1.04	0.01
Exp. ²⁹		(0.05 - 0.1)					1.6	
Exp. ⁶⁶		0.01					1.0	

mass was that of either $^{12}\text{C}^{16}\text{O}$ or $^{13}\text{C}^{18}\text{O}$. For the power spectra (Figure 2A) the isotope shift is 65 cm^{-1} (compared with 60 cm^{-1} from experiment)⁶⁵ whereas for the spectra from solving the 1D Schrödinger equation (Figures 2C) it is 74 cm^{-1} .

The FFCF for the INM and quantum bound state frequencies are reported in Figures 2B and C. At very short time ($\tau \sim 0.1 \text{ ps}$) a minimum in the FFCF is found which was not present in earlier work³⁵, probably due to the fact that the solute was frozen during the simulations. This feature is known from previous simulations⁶⁷ and related to the strength of the interaction between solvent and solute.^{34,35,59} These interactions are expected to decrease in going from CN^- , N_3^- , H_2O to the amide-I in N-methyl-acetamide due to the different strengths of the dipole and higher multipolar interactions with the environment. The two FFCFs show invariably multiexponential time decays. The parameters of the fits to Eqs. 2 and 3 are provided in Table 1 and compared with earlier work and experiment.^{29,66} The quality of the fits is good and the time decays τ_3 derived from them are consistent with the experiments which validate the computational model for using it in simulations of insulin monomer and dimer.

3.2 FFCF and Infrared Spectra for the Insulin Monomer

Next, the FFCFs and infrared spectra of -CO probes for insulin monomer in water were determined. For this, 2×10^5 snapshots, separated by 5 fs were analyzed from the MD simulations and the positionally resolved spectra were determined from analytically integrating the corresponding FFCFs, Eqs. 2 and 3. In all cases the FFCF was fitted either to Eq.2 or 3, depending on whether the minimum in the FFCF at short time ($\tau \leq 0.1$ ps) was present or not.

Typical FFCFs are reported in Figure 3A. The FFCFs for -CO probes at different positions along the polypeptide chain (see Figure 1) show that the local environmental dynamics differs. This is evident from changes in the curvature, the static components (as indicated by the value of $C(t = 5\text{ps})$), and the fact that some of the FFCFs have a minimum at short correlation times whereas others do not. The fact that some of the -CO probes exhibit a weak minimum whereas others do not suggests that the simulations are sensitive to such different probe/solvent interactions and will be discussed further below.

More specifically, the correlation functions for the -CO probes at residues A19, B6, B20, and B24 for the monomer together with the fits to the parametrized representations (Eqs. 2 and 3) are reported in Figure 3A. The initial amplitude $C(t = 0)$ ranges from 4 ps^{-2} for B6 to $\sim 8 \text{ ps}^{-2}$ for the other three residues considered. Similarly, the asymptotic value (here taken at 5 ps) differs for B6 compared with the other three residues. For A19, B20, and B24 the FFCF at longer times decays to a value close to zero whereas that for B6 remains at a finite value of $\sim 1 \text{ ps}^{-2}$, see inset of Figure 3A. Finally, at short times ($t \sim 0.1$ ps) the FFCFs for B6 (faint) and B20 (pronounced, see lower inset Figure 3A) exhibit a minimum, that for A19 displays a shoulder whereas that for B24 has no discernible feature. All FFCFs clearly contain multiple time scales and the fits to the empirical representations are of good quality. Overall, the FFCFs for -CO at different locations along the polypeptide chain show different,

clearly distinct features.

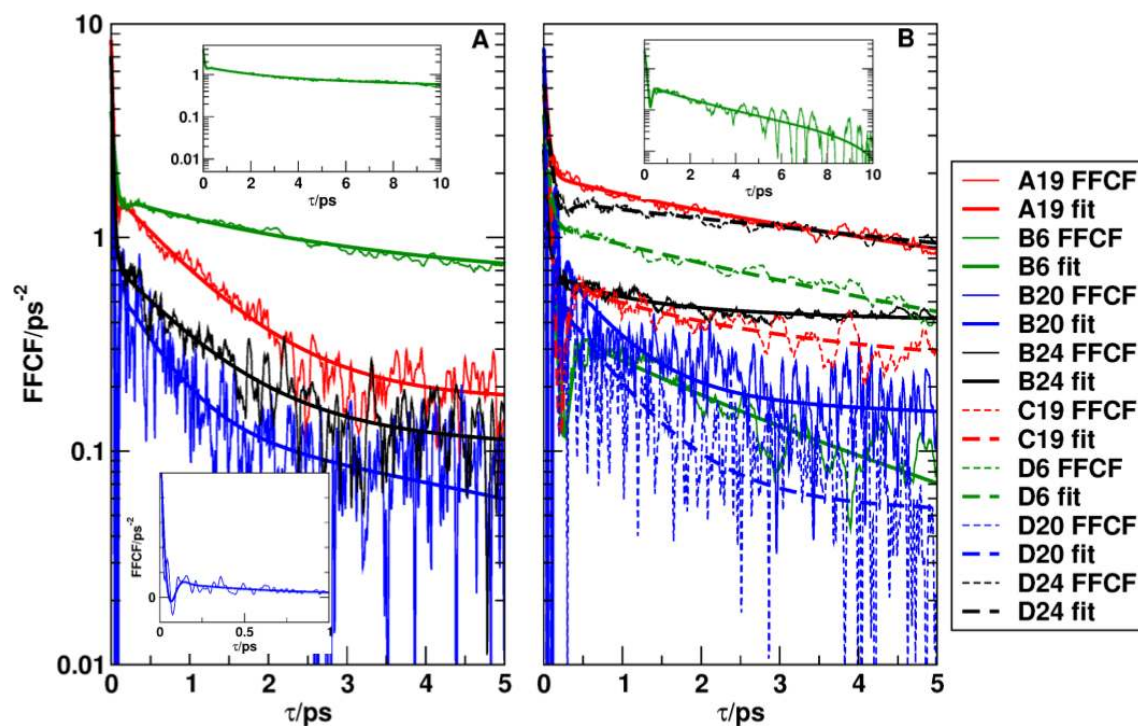
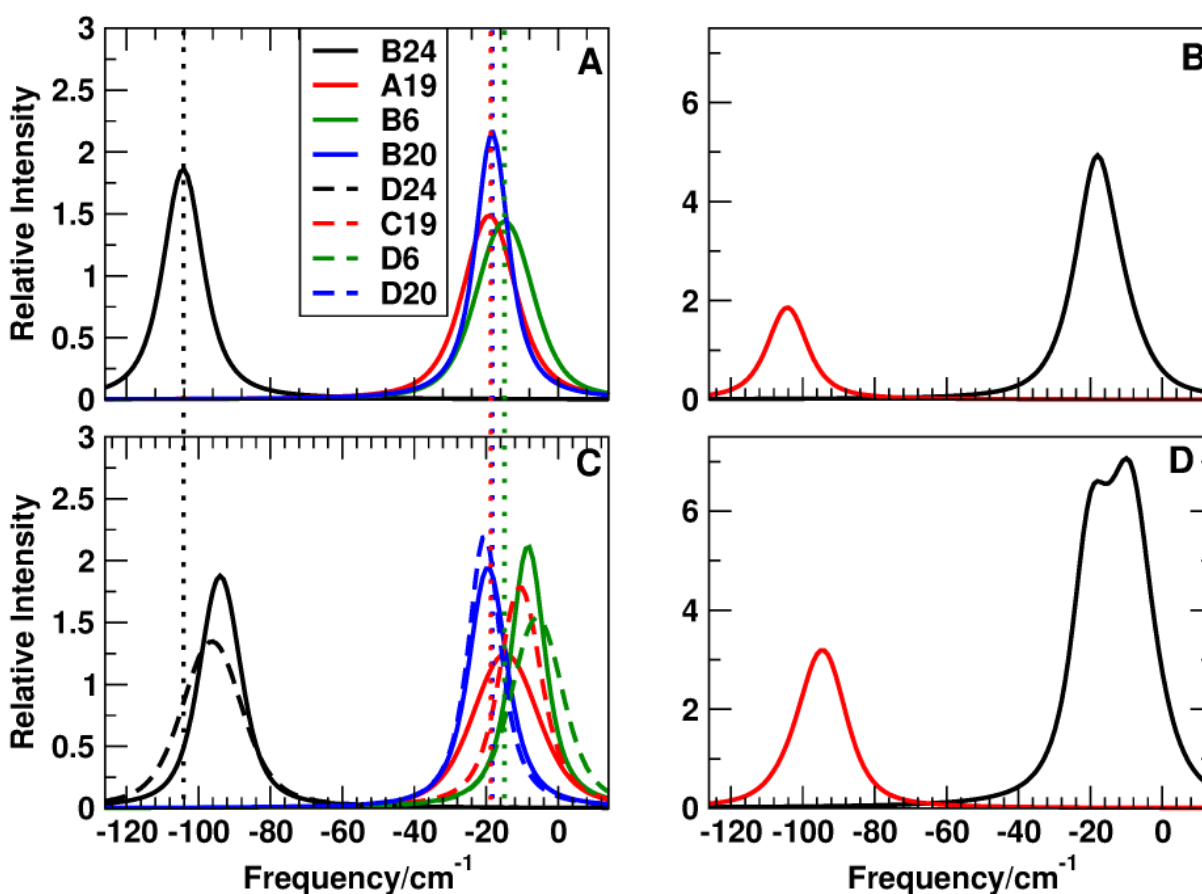


Figure 3: Frequency fluctuation correlation functions (FFCFs) for 3 selected $^{12}\text{C}^{16}\text{O}$ probes (A19, B6 and B20) and for $^{13}\text{C}^{18}\text{O}$ (B24) for the monomeric/dimeric forms. Solid thick lines for monomer I, dashed thick lines for monomer II (in the dimer). The raw data is shown as thin lines in the same color as the fits. Insets show the FFCF over longer times for the -CO at position B6 (blue traces) and for the first 1 ps (for B20, green trace and on linear y -scale) to highlight the minimum at short correlation time. See Figure 1 for location of the probes in the protein.

From the FFCFs the infrared lineshape can be directly determined by analytical integration. As for the FFCFs the 1-d infrared lineshapes differ depending on the position of the -CO label. Both, the maximum of the absorption and the full width at half maximum (FWHM) are influenced by the position of the probe along the protein chain, see Figure 4A. In the following all infrared frequencies are reported relative to the $^{12}\text{C}^{16}\text{O}$ frequency of NMA in solution from solving the 1-dimensional Schrödinger equation (at 1586 cm^{-1}). The band centers for all $^{12}\text{C}^{16}\text{O}$ in Figure 4A are shifted to the red by around 20 cm^{-1} . This is consistent with findings from experiments on $(\text{Ala})_3$ in water which find the -CO stretch at

1
 2
 3 $\sim 1600 \text{ cm}^{-1}$, shifted some 25 cm^{-1} to the red from the absorption for NMA.²⁹ The same
 4 observation is made when scanning along the amide-I normal mode, see Figure S2: The
 5 frequency distributions shift to the red compared to that of NMA by 10 to 20 cm^{-1} (for
 6 the two residues considered) but the overall distribution remains almost unchanged. The
 7 FWHMs for the lineshapes at positions B20 and B24 are comparable whereas those for B6
 8 and A19 are larger. These are also the residues for which the FFCF decayed more slowly or
 9 even remained at a finite static value at longer correlation times.
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19



20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48 Figure 4: 1D IR absorption spectra for 3 selected $^{12}\text{C}^{16}\text{O}$ bonds (A19, B6 and B20) and for
 49 $^{13}\text{C}^{18}\text{O}$ (B24) for monomeric (panels A and B) and dimeric (panels C and D) insulin. See
 50 Figure 1 for locations in the protein. Solid lines for monomer I and dashed lines for monomer
 51 II. Panels B and D report the superposition of all $^{12}\text{C}^{16}\text{O}$ lineshapes (red) compared with
 52 that of the single $^{13}\text{C}^{18}\text{O}$ probe (black).
 53
 54
 55
 56
 57
 58
 59
 60

3.3 FFCF and Infrared Spectra for the Insulin Dimer

The structural dynamics, FFCFs and 1D-infrared lineshapes were also determined for the homodimer using the same -CO probes as for the monomer. This provides information about relative shifts of the IR-absorption and changes in the FWHM. In addition, it is possible to address the question whether the dynamics of each of the monomers in the insulin dimer is symmetric, i.e. whether the crystallographic symmetry (C1) also extends to symmetry in the structural dynamics and the spectroscopic response.

Table 2: Fitting parameters of the FFCFs for $\omega(t)$ from solving the 1-d Schrödinger equation for $^{12}\text{C}^{16}\text{O}$ and $^{13}\text{C}^{18}\text{O}$ in monomeric and dimeric insulin for residues B24, A19, B6, B20, D24, C19, D6, D20. The data is fit to Eqs. 2 or 3.

	a_1 ps ⁻²	τ_1 ps	a_2 ps ⁻²	τ_2 ps	Δ_0 ps ⁻²	γ ps ⁻¹
B24 monomer	6.74	0.02	0.66	1.01	0.11	-
A19 monomer	7.04	0.02	1.47	0.99	0.17	-25.5
B6 monomer	2.40	0.03	0.92	2.82	0.59	0.2
B20 monomer	6.62	0.02	0.47	1.01	0.04	34.6
B24 dimer	1.74	0.05	0.26	1.40	0.41	-
A19 dimer	4.77	0.04	1.51	3.96	0.44	-
B6 dimer	2.00	0.10	0.35	3.15	0.00	8.3
B20 dimer	7.52	0.02	0.38	1.56	0.13	38.0
D24 dimer	3.54	0.06	1.41	11.02	0.05	-0.2
C19 dimer	3.62	0.08	0.36	3.15	0.22	-10.9
D6 dimer	2.49	0.06	0.97	4.09	0.16	-
D20 dimer	6.02	0.02	0.30	1.51	0.04	41.6

The relevant FFCFs for monomers A or B are reported in Figure 3B. It is found that all FFCFs change compared to those in the isolated monomer. Typically, the time scale for the long-time decay increases (see Table 2) for all probes considered. For B6 the minimum at short times is more pronounced for the dimer compared to the monomer which is also true for B20. For A19 the decay slows down considerably in the monomer and for B24, directly located at the dimerization interface, the static component also increases considerably.

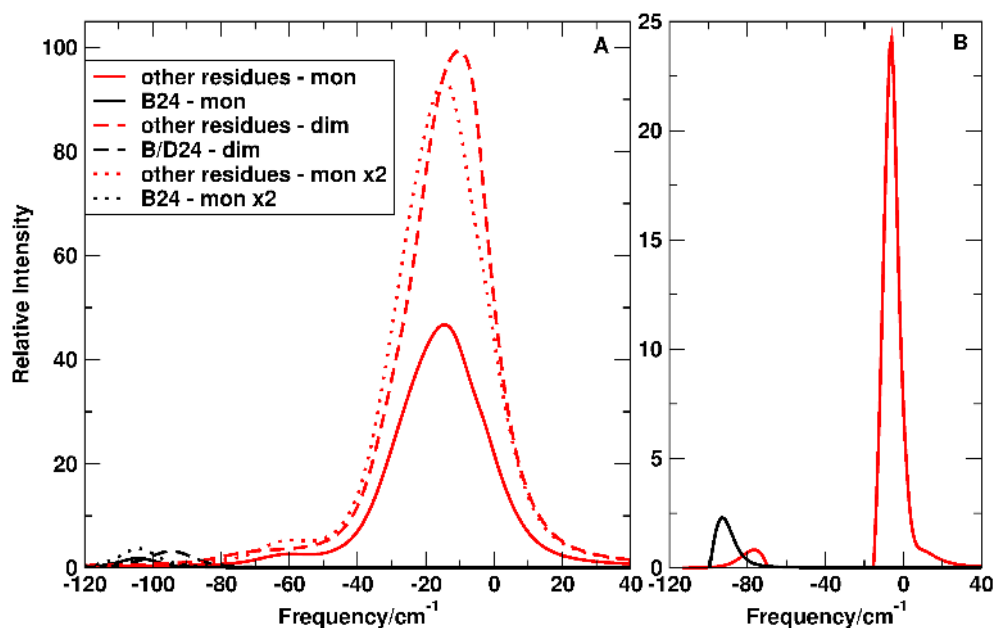


Figure 5: Panel A: The IR spectra of insulin monomer (solid line) and dimer (dashed line) from a superposition of all individual IR spectra. Also, twice the IR spectrum of the monomer is shown (dotted) which differs from the IR spectrum of the dimer and suggests an asymmetry between the two monomers in the dimer. Panel B: The difference spectrum between dimer and twice the monomer. Black traces are for B24 which is labelled.

These change in the FFCFs are also reflected in the IR spectroscopy. Both, the position of the absorption maxima and the FWHM differ in going from the monomer to the dimer. Specifically, $[\omega_{\max}, \text{FWHM}]$ (both in cm^{-1}) change as follows: for A19 $[-20, 16] \rightarrow [[-17, 16]$ and $[-12, 16]$, for B6 $[-12, 19] \rightarrow [[-9, 11]$ and $[-7, 16]]$, for B20 $[-18, 12] \rightarrow [[-19, 12]$ and $[-20, 10]]$, and for B24 $[-104, 14] \rightarrow [[-95, 13]$ and $[-94, 23]]$, see Figures 4A and C. The aggregate spectra for the three $^{12}\text{C}^{18}\text{O}$ and the $^{13}\text{C}^{18}\text{O}$ probe (at B24) for the monomer and the dimer are shown in Figures 4C and D. They also show that the IR-spectroscopy of the monomer and the dimer contain potentially valuable information about the aggregation state of the hormone, see also Figure 5, which establishes that the IR spectrum of the dimer differs from that of twice the monomer which should be observable experimentally.

It should be noted that the experimentally measured infrared spectrum of insulin dimer finds

1
2
3 the amide-I bands at around 1640 cm^{-1} , i.e. blue shifted relative to that found in NMA.⁶⁸
4
5 However, these experiments are carried out at low pH including DCl, NaCl and 20 % EtOD
6
7 as cosolvent. Conversely, for experiments with (Ala)₃ in water the amide-I absorptions shift
8
9 to the red by up to 20 cm^{-1} .²⁶
10

11
12
13 In summary, the dynamics and infrared spectroscopy of individual -CO probes along the
14
15 polypeptide changes is sensitive to the aggregation state of the hormone. In a next step,
16
17 these changes are discussed in the context of the overall structural dynamics of insulin
18
19 monomer and dimer.
20

21 22 23 **3.4 Structural Dynamics and Spectroscopy**

24
25
26 The dimerization interface of insulin involves residues B24 to B26 which are linked through
27
28 hydrogen bonds to residues D26 to D24. Considering this, the change in the FFCEs and
29
30 infrared spectra for the -CO probes of these residues upon dimerization are a primary target
31
32 for a structural interpretation of the spectroscopy. For the monomer the -CO stretches for
33
34 B25 and B26 superimpose (Figure 6A) whereas that for B24 is shifted to the red because
35
36 the heavy isotope was used in the simulation. Upon dimerization this changes. For the label
37
38 at B24 the bands shift by 8 to 10 cm^{-1} to the blue and the FWHM increase somewhat, see
39
40 Figure 6B. For the labels at positions B25 and B26 all maxima shift to the blue and the
41
42 linewidths increase. The frequency shift ranges from 5 to 20 cm^{-1} , consistent with findings
43
44 for (Ala)₃.²⁶ As for B24 the two symmetry-related positions in the X-ray structure split for
45
46 the structural dynamics and instead of 1 band a total of 4 bands emerge. Thus, dimeriza-
47
48 tion leads to noticeable changes in the IR-spectroscopy of the -CO probes depending on the
49
50 changes of their environment at the interface.
51

52
53
54 Another potentially interesting region that may be affected by dimerization is around the
55
56 turn involving residues B20 to B23 because these residues can interact with the end of the B-
57

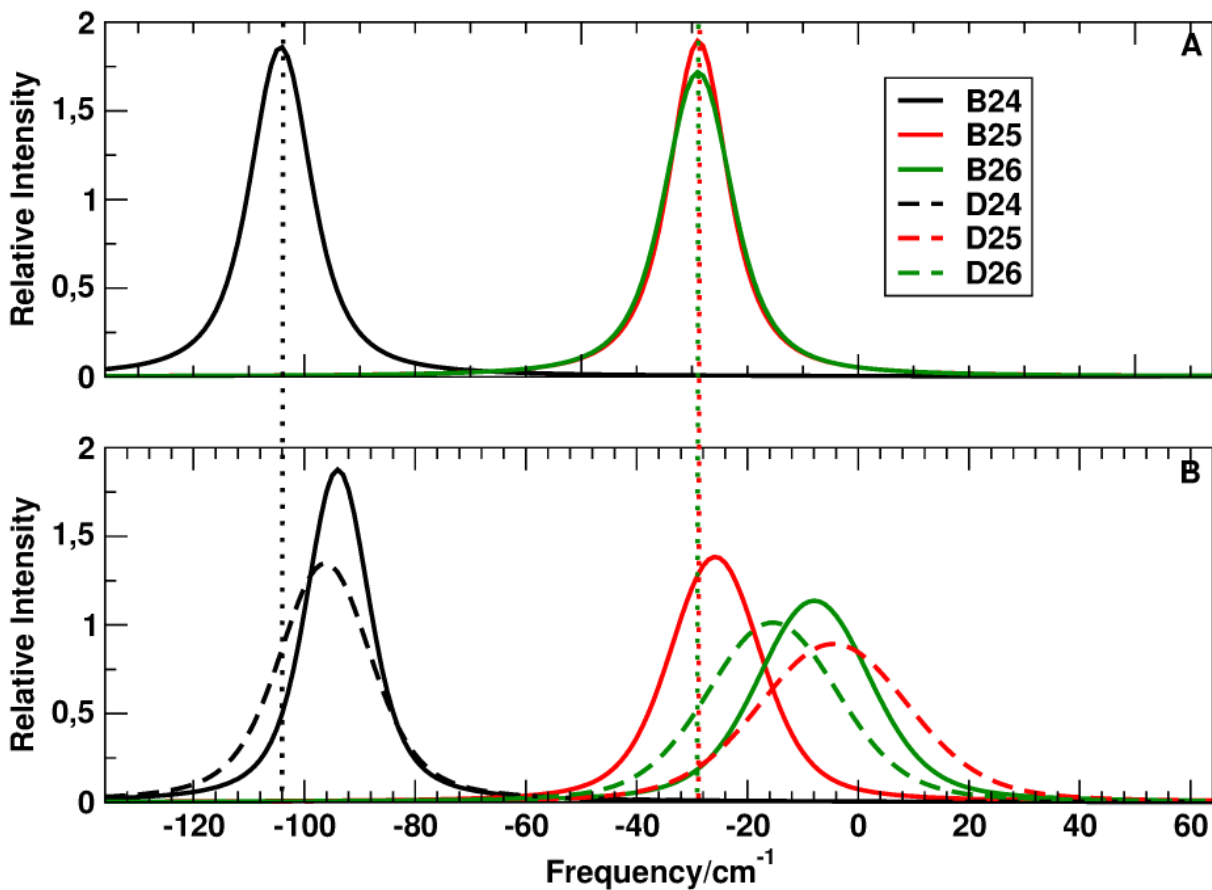
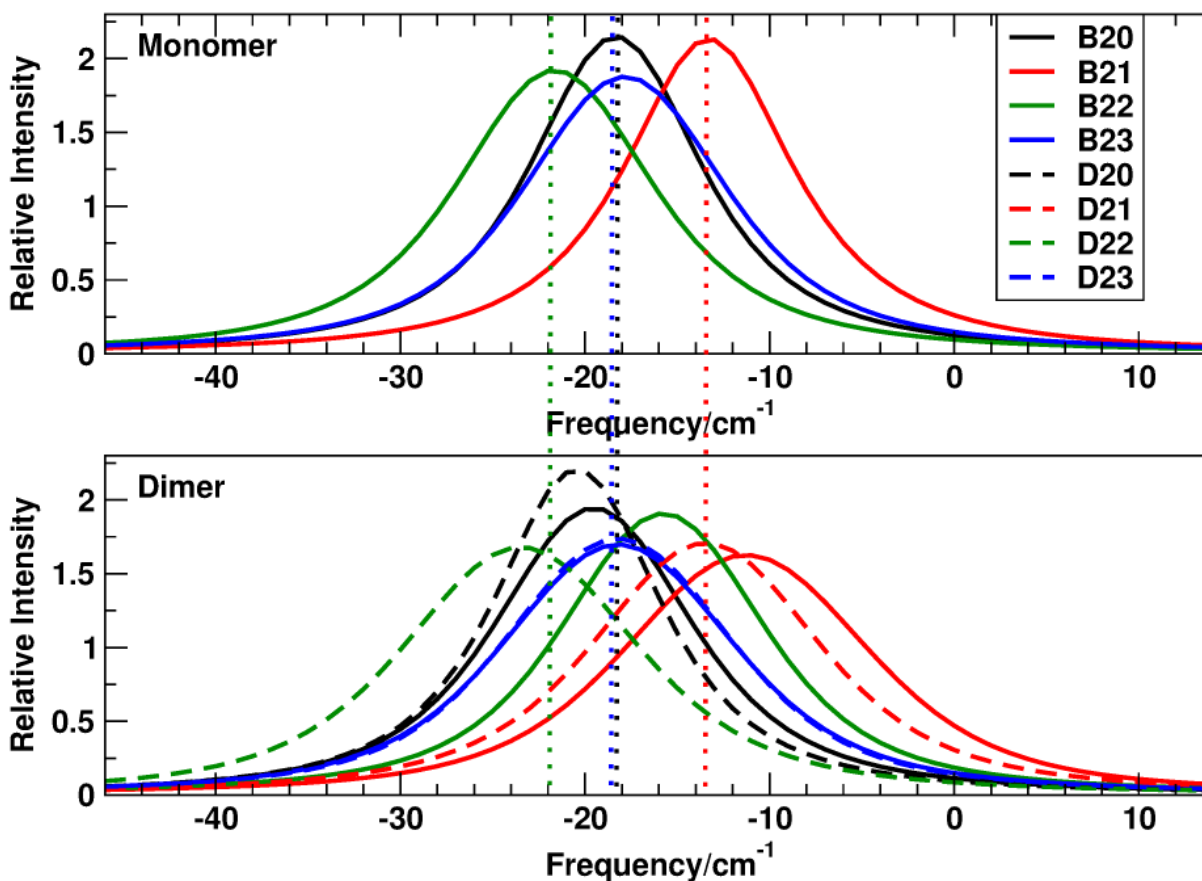


Figure 6: The IR spectra for residues B24 to B26 in the monomer (upper panel) and for the two monomers in the dimer (lower panel) for isotopically labelled B24.

1
2
3 chain (residues B27 onwards), see Figure 1. These IR spectra are shown in Figure 7. Again,
4
5 it is found that the spectra of the monomer and dimer do not overlap, see Figures 7A and
6
7 B. In addition, the IR spectra for the symmetry-related pairs in the B- and D-chains do not
8
9 superimpose which suggests that the structural dynamics of the dimer is not symmetric for
10
11 the two monomers in the homo-dimer. However, the effects of dimerization on the position
12
13 of the frequency maxima are considerably smaller ($\pm 5 \text{ cm}^{-1}$) than those for residues B24 to
14
15 B26 which are directly involved in dimerization through H-bonding for which they span ± 25
16
17 cm^{-1} .



49 Figure 7: The IR spectra for residues B20 to B23 in the monomer (upper panel) and for the
50 two monomers in the dimer (lower panel).
51
52
53
54

55 Given that the FFCFs and 1d-infrared response derived from it provide valuable informa-
56
57
58
59
60

tion about the environmental dynamics, FFCFs for all -CO labels in the monomer and both monomers in the insulin dimer were computed and characterized. Because the spectroscopic response will be determined by both, the water and the protein dynamics around the -CO label, it is expected that there are not particularly simple correlations between structure and infrared spectroscopy. Most absorption maxima of the $^{12}\text{C}^{16}\text{O}$ vibrations cover a range of 50 to 60 cm^{-1} compared with ~ 70 to 80 cm^{-1} from experiment.²⁵

In the following typical determinants that affect the spectroscopy around a probe are considered. First, the average number of water molecules n_W within a radius of 4 Å of the label is considered depending on the position of the -CO along the polypeptide chain, see Figure 8. It is found that with increasing average hydration $\langle n_W \rangle$ the infrared absorption shifts gradually to the red. This is consistent with a typical red shift of a -CO group upon interaction with the hydrogen atoms of the water molecules. However, the correlation is not particularly strong. In Figure 8A a linear correlation determined from all data (red) and a line (blue) connecting the average $(\langle n_w \rangle, \langle \omega \rangle)$ for the lowest ($n_W \sim 0$) and largest ($n_W \sim 4$) degree of hydration both illustrate this. Hence, the overall frequency shifts can be rationalized based on local interactions with hydration waters. The water-exposure of residues F24, F25, and Y26 changes appreciably between the monomeric and dimeric form, as expected. They are $n_w \geq 3$ for the monomer and reduce to $n_w \sim 1$ for the dimer. The fact that water can penetrate into the dimerization interface has recently been found in simulations for WT and mutant insulins at position B24.¹³

Next, the value $C(t = 0)$ of the FFCF is compared with the average number of water molecules. Again, there is a weak linear correlation (Figure 9) between the two quantities which indicates that with increasing hydration the amplitude $C(t = 0)$ increases. This (weak) correlation can also be understood in a qualitative fashion because the amplitude of the FFCF is related to the interaction between the -CO probe and the environment which

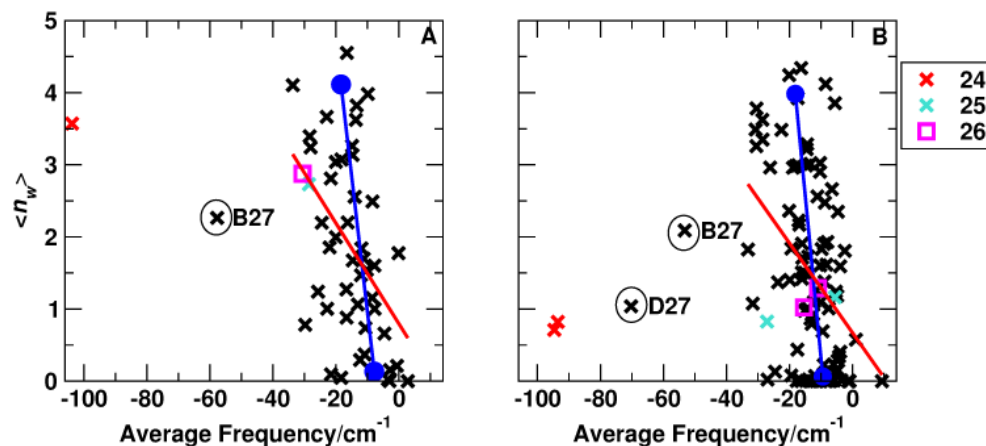


Figure 8: The maximum of the IR spectrum *vs.* the average number of water molecules within a 4 Å radius of the O atom of the amide bond for the monomer. Residues B24, B25, B26 and the outliers are labelled specifically. The red line is the linear correlation for all data except for B24 and the outliers. The blue line connects the centers of gravity (blue dots) for $n_w \sim 0$ and $n_w \sim 4$. There is no ready structural explanation for the outlier at position 27.

leads to a shift of the average stretching frequency to the red, see Figure 8

Several other determinants of the spectra were compared, e.g. the maximum of the absorption band with the fluctuation of the frequency itself or with the average residence time of the water molecules (see Figure S3). However, these were not found to be related even in a qualitative fashion. In summary, there is no single, simple structural feature of the conformational dynamics that can be related in a straightforward and unambiguous manner with the frequency shift of the -CO oscillator. This is also not to be expected given the demanding environment of each of the oscillators that changes on the picosecond time scale.

A final question concerns the origin of the frequency shifts which was assessed in the following way for a typical -CO label that is surrounded by both, water and neighboring protein residues. The -CO group of the LeuB17 residue in the B16-B17-B18 (Tyr-Leu-Val) tripeptide has been chosen for this as it is surrounded by both, water and protein environments. Fifty snapshots were extracted from the MD simulations of the insulin monomer in water and pre-

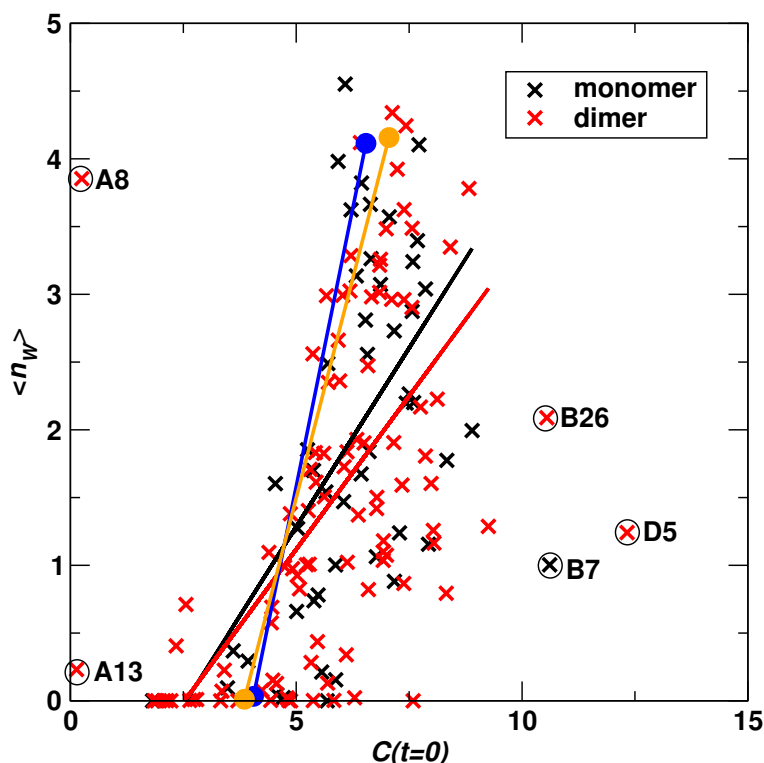


Figure 9: The value of $C(t = 0)$ vs. the average number of waters within a 4 \AA radius of the O atom of the amide bond. The black and red solid lines are the linear correlations for all data except the outliers for monomer and dimer, respectively. The blue (monomer) and yellow (dimer) solid lines connecting the filled circles (center of gravity) for $n_w \sim 0$ and $n_w \sim 4$. There is no obvious explanation for the outliers (in circles and labelled).

pared in the following way. The tripeptide and all surrounding water molecules and amino acid residues within 5 \AA of the $-\text{CO}$ label of LeuB17 were selected (typically 100 atoms) for each snapshot. For each snapshot the 1-dimensional potential energy surface was scanned along the $-\text{CO}$ normal mode at the B3LYP/6-31+G(d,p) level of theory using Gaussian09⁶⁹ and the transition frequency was determined by solving the 1-dimensional Schrödinger equation⁷⁰ on the *ab initio* PES. This was repeated for the systems in which (i) all water molecules, (ii) all residues surrounding the tripeptide and (iii) both, water molecules and surrounding residues were removed (“gas phase”). Thus, in total 4 frequency distributions were obtained and the frequency shifts induced by the water, amino acid residues and both were computed with respect to the gas phase frequency of the tripeptide for each snapshot.

The results in Figure 10 show that the frequency distribution of the full system (tripeptide + water + closest protein residues, green) is shifted by $\sim -55 \text{ cm}^{-1}$ to the red relative to the gas phase tripeptide frequencies. The shift induced by the closest protein residues (red) is only $\sim -10 \text{ cm}^{-1}$ and is relatively narrow ($\pm 10 \text{ cm}^{-1}$) whereas that due to the water shell (blue) amounts to $\sim -50 \text{ cm}^{-1}$ and is comparable in width to the green distribution. This suggests that most of the spectroscopic response for this -CO probe is due to the water shell.

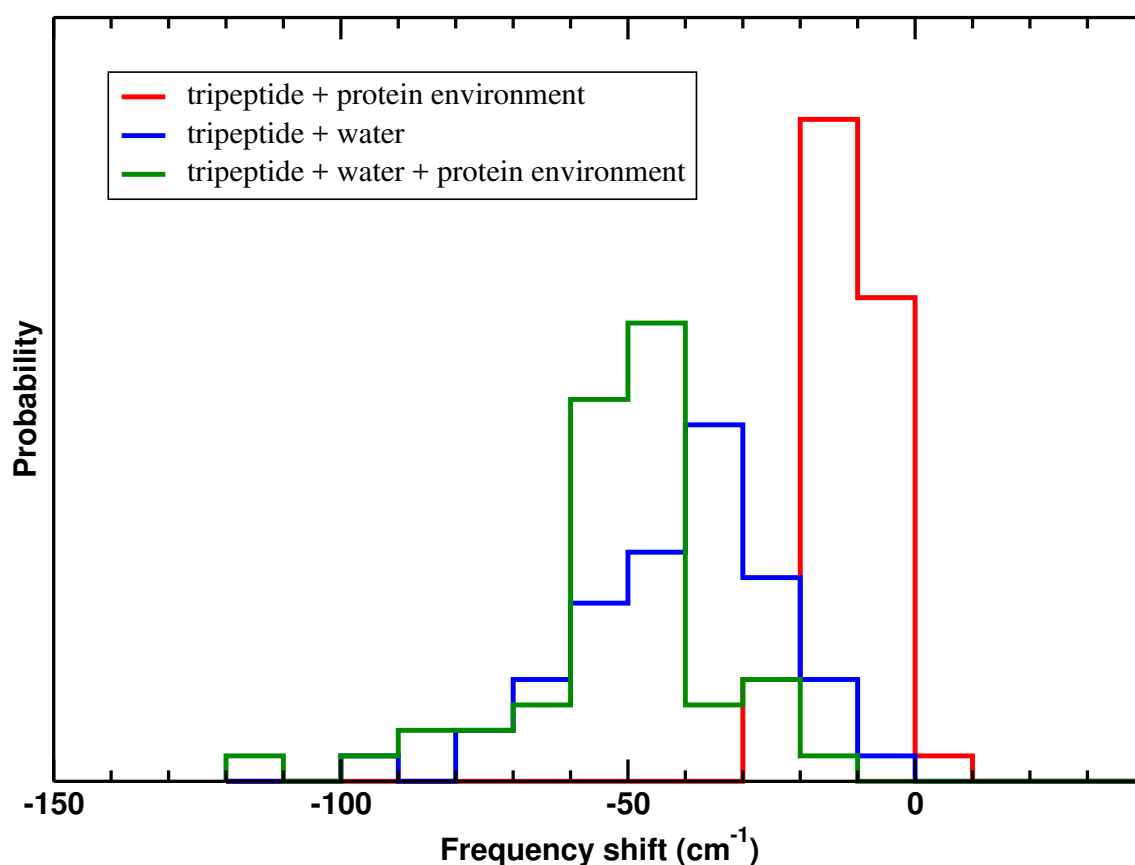


Figure 10: Distribution of frequency shifts for the -CO stretch mode of LeuB17 in B16-B17-B18 obtained from 50 snapshots for the tripeptide is either surrounded by amino acid residues (red) or water molecules (blue) and both the amino acid residues and water (green). The zero of frequency is the gas phase frequency of the tripeptide for each snapshot (see text for details).

This is a quantitative assessment of the influence of the water and the protein environment

1
2
3 on the -CO transition frequency. It demonstrates that for water-exposed residues the major
4 effect arises from the coordinating water molecules and a considerably smaller part from the
5 protein environment.
6
7
8
9

10 11 12 13 4 Discussion and Conclusions

14
15
16 Atomistic simulations using a validated force field for the spectroscopy of -CO probes for
17 insulin monomer and dimer establish that the -CO stretch is sensitive to both, the environ-
18 mental dynamics and the aggregation state of the hormone. The -CO probe frequencies span
19 a range of ± 20 cm^{-1} for the monomer and ± 25 cm^{-1} for the dimer, see Figure 8. Such a
20 magnitude is consistent with previous experiments⁷¹ and simulations.⁷²
21
22
23
24
25
26
27

28
29 The dimerization of insulin has been investigated using a range of experimental techniques.
30 Using UV-vis spectroscopy dimerization free energies ranging from -8.2 kcal/mol to -6.6
31 kcal/mol were found^{6,73} which compare with -5.4 kcal/mol from analysis of amide-I two-
32 dimensional spectroscopy (note the different definition of the reference state in Refs.^{6,73}
33 compared to Ref.²⁴).²⁴ Thus, depending on the technique used, the results can be quite
34 different. Furthermore, a pronounced sensitivity of the stability of the dimer to addition
35 of cosolvent (e.g. ethanol) has been reported.²⁴ One of the challenges in comparing exper-
36 imental and computational studies is in the solvent conditions that typically are used in
37 spectroscopic studies where ethanol, HCl and NaCl are usually added.
38
39
40
41
42
43
44
45
46
47

48
49 It has been previously suggested⁷¹ and more recently verified⁷² that instead of carrying out
50 a ligand-binding experiment it is possible to use infrared spectroscopy to characterize the
51 binding strength of a ligand towards its target protein. Based on this finding the present
52 work sets out to characterize the differences in the spectroscopic response of the insulin
53
54
55
56
57
58
59
60

1
2
3 monomer and dimer upon aggregation in the amide-I region. This may open up an avenue
4 to complement a thermodynamic experiment with a potentially more convenient spectro-
5 scopic characterization of the aggregation state.
6
7
8
9

10
11 With regards to dimerization it is demonstrated that the dynamics and spectroscopy of the
12 -CO probes involved in the protein-protein contact (B24 to B26 and D24 to D26) exhibit char-
13 acteristic changes. In particular, spectroscopic signatures of symmetry-related -CO probes
14 in the X-ray structure split into two distinct bands which indicates that the structure of the
15 insulin dimer is asymmetric when the two monomers are considered. This is also consistent
16 with the X-ray structures which explicitly note on the asymmetry of the monomer structures
17 within the dimer.^{1,74} The rather large number of -CO oscillators (49 for the monomer, 98 for
18 the dimer) potentially leads to a dilution of the differential spectroscopic signatures. But it
19 may still be sufficient to distinguish monomeric from dimeric insulin in solution. If position-
20 ally resolved information is sought, isotopically substituted -CO labels will be useful.²⁵
21
22
23
24
25
26
27
28
29
30
31

32
33 For most of the analysis, the local -CO stretch mode was sampled which led us to report
34 frequency shifts relative to the gas phase absorption in NMA. However, it would also be
35 possible to determine the 1-dimensional potential by scanning along the normal mode corre-
36 sponding to the -CO stretch (see further above for a discussion of NMA). In order to quantify
37 the expected changes in the absolute frequencies, this was done for 10^5 snapshots for the
38 -CO label at HisB5 and TyrB26, see Figure S2. It is found that both frequencies shift by
39 $\sim 60 \text{ cm}^{-1}$ to the blue into the region where typically absorptions of insulin in solution are
40 found whereas the widths of the distributions, and hence the lineshapes, remain the same.
41 Thus, in the future, it may be advantageous to determine the energy along the normal mode
42 of the -CO stretch instead of scanning the -CO local mode.
43
44
45
46
47
48
49
50
51
52
53

54
55 Atomistic simulations using validated energy functions together with state-of-the art ex-
56
57
58
59
60

1
2
3 experiments provide a unique opportunity to relate the spectroscopic signatures with the un-
4 derlying structural dynamics with positional resolution.⁷⁵ This together with free energy
5 simulations¹³ opens up the possibility to characterize a broader range of modified insulins
6 with respect to their thermodynamic stability of the dimers which is an essential determinant
7 of the physiological mode-of-action of the hormone.
8
9
10
11
12
13
14
15
16

17 Supporting Information Available

18
19
20 Power spectra and FFCFs for NMA from vibrational mode analysis, Frequency distributions
21 obtained from scanning along the local and normal mode of -CO for TyrB26 and HisB5,
22 Frequency fluctuation, fluctuation of the number of water molecules and average residence
23 time of the water molecule for insulin monomer and dimer. This material is available free
24 of charge via the Internet at <http://pubs.acs.org/>.
25
26
27
28
29
30
31
32

33 Acknowledgement

34
35
36 This work was supported by the Swiss National Science Foundation grant 200021-117810,
37 the NCCR MUST, and the University of Basel. We thank Prof. J. L. Skinner for comments
38 on part of this manuscript and Prof. A. Tokmakoff for discussions on the solvent-sensitivity
39 of characterizing the monomer-dimer equilibrium.
40
41
42
43
44
45
46
47

48 References

- 49
50
51 (1) Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D.
52 M. C.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D.; Sakabe, K. et al. The Structure
53
54
55
56
57
58
59
60

- 1
2
3 of 2Zn Pig Insulin Crystals at 1.5 Å Resolution. *Philos. Trans. R. Soc. London, Ser. B*
4 **1988**, *319*, 369–456.
5
6
7
- 8 (2) Jørgensen, A. M. M.; Olsen, H. B.; Balschmidt, P.; Led, J. J. Solution Structure of the
9 Superactive Monomeric Des-[Phe (B25)] Human Insulin Mutant: Elucidation of the
10 Structural Basis for the Monomerization of Des-[Phe (B25)] Insulin and the Dimeriza-
11 tion of Native Insulin. *J. Mol. Biol.* **1996**, *257*, 684–699.
12
13
14
15
- 16 (3) Brems, D. N.; Alter, L. A.; Beckage, M. J.; Chance, R. E.; DiMarchi, R. D.; Green, L. K.;
17 Long, H. B.; Pekar, A. H.; Shields, J. E.; Frank, B. H. Altering the Association Prop-
18 erties of Insulin by Amino Acid Replacement. *Protein Eng.* **1992**, *5*, 527–533.
19
20
21
22
- 23 (4) Zoete, V.; Meuwly, M.; Karplus, M. Study of the Insulin Dimerization: Binding Free
24 Energy Calculations and Per-Residue Free Energy Decomposition. *Proteins: Struct.*
25 *Funct. Bioinf.* **2005**, *61*, 79–93.
26
27
28
29
- 30 (5) Pocker, Y.; Biswas, S. B. Self-Association of Insulin and the Role of Hydrophobic
31 Bonding: a Thermodynamic Model of Insulin Dimerization. *Biochemistry* **1981**, *20*,
32 4354–4361.
33
34
35
36
- 37 (6) Strazza, S.; Hunter, R.; Walker, E.; Darnall, D. W. The Thermodynamics of Bovine
38 and Porcine Insulin and Proinsulin Association Determined by Concentration Difference
39 Spectroscopy. *Arch. Biochem. Biophys.* **1985**, *238*, 30–42.
40
41
42
43
- 44 (7) Hua, Q. X.; Shoelson, S. E.; Kochoyan, M.; Weiss, M. A. Receptor Binding Redefined
45 by a Structural Switch in a Mutant Human Insulin. *Nature* **1991**, *354*, 238–241.
46
47
48
- 49 (8) Chen, H.; Shi, M.; Guo, Z.-Y.; Tang, Y.-H.; Qiao, Z.-S.; Liang, Z.-H.; Feng, Y.-M. Four
50 New Monomeric Insulins Obtained by Alanine Scanning the Dimer-Forming Surface of
51 the Insulin Molecule. *Protein Eng.* **2000**, *13*, 779–782.
52
53
54
55
56
57
58
59
60

- 1
2
3 (9) DeFelippis, M. R.; Chance, R. E.; Frank, B. H. Insulin Self-Association and the Rela-
4 tionship To Pharmacokinetics and Pharmacodynamics. *Crit. Rev. Ther. Drug Carrier*
5 *Syst.* **2001**, *18*, 201.
6
7
8
9
10 (10) Antolíková, E.; Žáková, L.; Turkenburg, J. P.; Watson, C. J.; Hančlová, I.; Šanda, M.;
11 Cooper, A.; Kraus, T.; Brzozowski, A. M.; Jiráček, J. Non-Equivalent Role of Inter-And
12 Intramolecular Hydrogen Bonds in the Insulin Dimer Interface. *J. Bio. Chem.* **2011**,
13 *286*, 36968–36977.
14
15
16
17
18 (11) Jiracek, J.; Zakova, L.; Antolikova, E.; Watson, C. J.; Turkenburg, J. P.; Dodson, G. G.;
19 Brzozowski, A. M. Implications for the Active Form of Human Insulin Based on the
20 Structural Convergence of Highly Active Hormone Analogues. *Proc. Natl. Acad. Sci.*
21 *USA* **2010**, *107*, 1966–1970.
22
23
24
25
26
27 (12) Zoete, V.; Meuwly, M.; Karplus, M. A Comparison of the Dynamic Behavior of
28 Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active
29 Monomer. *J. Mol. Biol.* **2004**, *342*, 913–929.
30
31
32
33 (13) Raghunathan, S.; El Hage, K.; Desmond, J. L.; Zhang, L.; Meuwly, M. The Role of
34 Water in the Stability of Wild-type and Mutant Insulin Dimers. *J. Phys. Chem. B*
35 **2018**, *122*, 7038–7048.
36
37
38
39
40 (14) Kristensen, C.; Kjeldsen, T.; Wiberg, F. C.; Schäffer, L.; Hach, M.; Havelund, S.;
41 Bass, J.; Steiner, D. F.; Andersen, A. S. Alanine Scanning Mutagenesis of Insulin. *J.*
42 *Bio. Chem.* **1997**, *272*, 12978–12983.
43
44
45
46
47 (15) Derewenda, U.; Derewenda, Z.; Dodson, E.; Dodson, G.; Bing, X.; Markussen, J. X-Ray
48 Analysis of the Single Chain B29-A1 Peptide-Linked Insulin Molecule: A Completely
49 Inactive Analogue. *J. Mol. Biol.* **1991**, *220*, 425–433.
50
51
52
53 (16) Vashisth, H.; Abrams, C. F. All-Atom Structural Models of Insulin Binding To the
54
55
56
57
58
59
60

- 1
2
3 Insulin Receptor in the Presence of a Tandem Hormone-Binding Element. *Proteins*
4 *Struct., Funct. Bioinf.* **2013**, *81*, 1017–1030.
5
6
7
8 (17) Mirmira, R.; Tager, H. Role of the Phenylalanine B24 Side Chain in Directing Insulin
9 Interaction with Its Receptor. Importance of Main Chain Conformation. *J. Bio. Chem.*
10 **1989**, *264*, 6349–6354.
11
12
13
14 (18) Nakagawa, S. H.; Tager, H. S. Importance of Aliphatic Side-Chain Structure at Positions
15 2 and 3 of the Insulin A Chain in Insulin-Receptor Interactions. *Biochemistry* **1992**,
16 *31*, 3204–3214.
17
18
19
20
21 (19) Cutfield, J.; Cutfield, S.; Dodson, E.; Dodson, G.; Hodgkin, D.; Reynolds, C. Evidence
22 Concerning Insulin Activity from the Structure of a Cross-Linked Derivative. *Hoppe-*
23 *Seyler's Z. Physiol. Chem.* **1981**, *362*, 755–762.
24
25
26
27
28 (20) Ludvigsen, S.; Olsen, H. B.; Kaarsholm, N. C. A Structural Switch in a Mutant Insulin
29 Exposes Key Residues for Receptor Binding. *J. Mol. Biol.* **1998**, *279*, 1–7.
30
31
32
33 (21) Nakagawa, S. H.; Tager, H. Role of the Phenylalanine B25 Side Chain in Directing
34 Insulin Interaction with Its Receptor. Steric and Conformational Effects. *J. Bio. Chem.*
35 **1986**, *261*, 7332–7341.
36
37
38
39
40 (22) Shoelson, S. E.; Lu, Z. X.; Parlautan, L.; Lynch, C. S.; Weiss, M. A. Mutations at
41 the Dimer, Hexamer, and Receptor-Binding Surfaces of Insulin Independently Affect
42 Insulin-Insulin and Insulin-Receptor Interactions. *Biochemistry* **1992**, *31*, 1757–1767.
43
44
45
46
47 (23) Dong, J.; Wan, Z.; Popov, M.; Carey, P. R.; Weiss, M. A. Insulin Assembly Damps
48 Conformational Fluctuations: Raman Analysis of Amide I Linewidths in Native States
49 and Fibrils. *J. Mol. Biol.* **2003**, *330*, 431–442.
50
51
52
53
54 (24) Ganim, Z.; Jones, K. C.; Tokmakoff, A. Insulin Dimer Dissociation and Unfolding
55
56
57
58
59
60

- 1
2
3 Revealed by Amide I Two-Dimensional Infrared Spectroscopy. *Phys. Chem. Chem.*
4 *Phys.* **2010**, *12*, 3579–3588.
5
6
7
- 8 (25) Dhayalan, B.; Fitzpatrick, A.; Mandal, K.; Whittaker, J.; Weiss, M. A.; Tokmakoff, A.;
9 Kent, S. B. H. Efficient Total Chemical Synthesis of $^{13}\text{C}=^{18}\text{O}$ Isotopomers of Human
10 Insulin for Isotope-Edited FTIR. *ChemBioChem* **2016**, *17*, 415–420.
11
12
13
- 14 (26) Woutersen, S.; Hamm, P. Structure Determination of Trialanine in Water Using Polar-
15 ization Sensitive Two-Dimensional Vibrational Spectroscopy. *J. Phys. Chem. B* **2000**,
16 *104*, 11316–11320.
17
18
19
- 20 (27) Asplund, M.; Zanni, M.; Hochstrasser, R. Two-Dimensional Infrared Spectroscopy of
21 Peptides by Phase-Controlled Femtosecond Vibrational Photon Echoes. *Proc. Natl.*
22 *Acad. Sci. USA* **2000**, *97*, 8219–8224.
23
24
25
- 26 (28) Nizkorodov, S. A.; Dopfer, O.; Ruchti, T.; Meuwly, M.; Maier, J. P.; Bieske, E. J. Size
27 Effects in Cluster Infrared Spectra: the ν_1 Band of $\text{Ar}_n\text{-HCO}^+$ ($n = 1\text{--}13$). *J. Phys.*
28 *Chem.* **1995**, *99*, 17118–17129.
29
30
31
- 32 (29) Woutersen, S.; Pfister, R.; Hamm, P.; Mu, Y.; Kosov, D.; Stock, G. Peptide Conforma-
33 tional Heterogeneity Revealed from Nonlinear Vibrational Spectroscopy and Molecular-
34 Dynamics Simulations. *J. Chem. Phys.* **2002**, *117*, 6833–6840.
35
36
37
- 38 (30) Corcelli, S.; Lawrence, C.; Skinner, J. Combined Electronic Structure/molecular Dy-
39 namics Approach for Ultrafast Infrared Spectroscopy of Dilute HOD in Liquid H_2O and
40 D_2O . *J. Chem. Phys.* **2004**, *120*, 8107–8117.
41
42
43
- 44 (31) Gruenbaum, S. M.; Tainter, C. J.; Shi, L.; Ni, Y.; Skinner, J. L. Robustness of Fre-
45 quency, Transition Dipole, and Coupling Maps for Water Vibrational Spectroscopy. *J.*
46 *Chem. Theory Comput.* **2013**, *9*, 3109–3117.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (32) Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. Development and Validation
4 of Transferable Amide I Vibrational Frequency Maps for Peptides. *J. Phys. Chem. B*
5 **2011**, *115*, 3713–3724.
6
7
8
9
10 (33) Hayashi, T.; Jansen, T.; Zhuang, W.; Mukamel, S. Collective Solvent Coordinates for
11 the Infrared Spectrum of HOD in D2O Based on an Ab Initio Electrostatic Map. *J.*
12 *Phys. Chem. A* **2005**, *109*, 64–82.
13
14
15
16 (34) Lee, M. W.; Carr, J. K.; Göllner, M.; Hamm, P.; Meuwly, M. 2D IR Spectra of Cyanide
17 in Water Investigated by Molecular Dynamics Simulations. *J. Chem. Phys.* **2013**, *139*,
18 054506.
19
20
21
22
23 (35) Cazade, P.-A.; Bereau, T.; Meuwly, M. Computational Two-Dimensional Infrared Spec-
24 troscopy without Maps: N-Methylacetamide in Water. *J. Phys. Chem. B* **2014**, *118*,
25 8135–8147.
26
27
28
29
30 (36) Salehi, M.; Koner, D.; Meuwly, M. Vibrational Spectroscopy of N_3^- in the Gas- and
31 Condensed-Phase. *J. Phys. Chem. B* **2019**, *123*, 3282–3290.
32
33
34
35 (37) Ho, T.-S.; Rabitz, R. A General Method for Constructing Multidimensional Molecular
36 Potential Energy Surfaces from ab Initio Calculations. *J. Chem. Phys.* **1996**, *104*,
37 2584–2597.
38
39
40
41
42 (38) Unke, O. T.; Meuwly, M. Toolkit for the Construction of Reproducing Kernel-Based
43 Representations of Data: Application to Multidimensional Potential Energy Surfaces.
44 *J. Chem. Inf. Model.* **2017**, *57*, 1923–1931.
45
46
47
48
49 (39) Bloem, R.; Dijkstra, A. G.; Jansen, T. L. C.; Knoester, J. Simulation of Vibrational
50 Energy Transfer in Two-Dimensional Infrared Spectroscopy of Amide I and Amide II
51 Modes in Solution. *J. Chem. Phys.* **2008**, *129*.
52
53
54
55
56
57
58
59
60

- 1
2
3 (40) Brooks, B. R.; Brooks III, C. L.; MacKerell Jr., A. D.; Nilsson, L.; Petrella, R. J.;
4 Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. et al. CHARMM: The
5 Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
6
7
8
9
10 (41) J. A. MacKerell, *et. al.*. All-Atom Empirical Potential for Molecular Modeling and
11 Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
12
13
14 (42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W. Comparison of
15 Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*,
16 926–935.
17
18
19
20
21 (43) van Gunsteren, W.; Berendsen, H. Algorithms for Macromolecular Dynamics and Con-
22 straint Dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.
23
24
25
26 (44) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation
27 Method for the Calculation of Equilibrium Constants for the Formation of Physical
28 Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*,
29 637–649.
30
31
32
33
34 (45) Nosé, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Meth-
35 ods. *J. Chem. Phys.* **1984**, *81*, 511–519.
36
37
38
39 (46) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys.*
40 *Rev. A* **1985**, *31*, 1695–1697.
41
42
43
44 (47) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Tem-
45 perature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
46
47
48
49 (48) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems.
50 *Mol. Phys.* **1983**, *50*, 1055–1076.
51
52
53
54 (49) Hairer, E.; Lubich, C.; Wanner, G. Geometric Numerical Integration Illustrated by the
55 Störmer/Verlet Method. *Acta Numerica* **2003**, *12*, 399–450.
56
57
58

- 1
2
3 (50) Steinbach, P. J.; Brooks, B. R. New Spherical-Cutoff Methods for Long-Range Forces
4 in Macromolecular Simulation. *J. Comp. Chem.* **1994**, *15*, 667–683.
5
6
7
8 (51) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald
9 Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
10
11
12 (52) Kramer, C.; Gedeck, P.; Meuwly, M. Atomic Multipoles: Electrostatic Potential Fit,
13 Local Reference Axis Systems and Conformational Dependence. *J. Comput. Chem.*
14 **2012**, *33*, 1673–1688.
15
16
17 (53) Bereau, T.; Kramer, C.; Meuwly, M. Leveraging Symmetries of Static Atomic Multipole
18 Electrostatics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*,
19 5450–5459.
20
21
22 (54) Cazade, P.-A.; Hedin, F.; Xu, Z.-H.; Meuwly, M. Vibrational Relaxation and Energy
23 Migration of N-Methylacetamide in Water: The Role of Nonbonded Interactions. *J.*
24 *Phys. Chem. B* **2015**, *119*, 3112–3122.
25
26
27 (55) Kubelka, J.; Keiderling, T. Ab Initio Calculation of Amide Carbonyl Stretch Vibrational
28 Frequencies in Solution with Modified Basis Sets. 1. N-Methyl Acetamide. *J. Phys.*
29 *Chem. A* **2001**, *105*, 10922–10928.
30
31
32 (56) Roy, R. J. Level 8.2: A Computer Program for Solving the Radial Schrödinger Equation
33 for Bound and Quasibound Levels. *University of Waterloo Chemical Physics Research*
34 *Report CP-663* **2014**,
35
36
37 (57) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge
38 University Press: New York, 2011.
39
40
41 (58) Hamm, P.; Lim, M.; Hochstrasser, R. Structure of the Amide I Band of Peptides Mea-
42 sured by Femtosecond Nonlinear-Infrared Spectroscopy. *J. Phys. Chem. B* **1998**, *102*,
43 6123–6138.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (59) Moller, K.; Rey, R.; Hynes, J. Hydrogen Bond Dynamics in Water and Ultrafast Infrared
4 Spectroscopy: A Theoretical Study. *J. Phys. Chem. A* **2004**, *108*, 1275–1289.
5
6
7
8 (60) Jansen, T.; Knoester, J. A Transferable Electrostatic Map for Solvation Effects on
9 Amide I Vibrations and Its Application To Linear and Two-Dimensional Spectroscopy.
10 *J. Chem. Phys.* **2006**, *124*.
11
12
13
14 (61) Mayne, L. C.; Hudson, B. Resonance Raman Spectroscopy of N-methylacetamide:
15 Overtones and Combinations of the Carbon-Nitrogen Stretch (amide II') and Effect
16 of Solvation on the Carbon-Oxygen Double-Bond Stretch (amide I) Intensity. *J. Phys.*
17 *Chem.* **1991**, *95*, 2962–2967.
18
19
20
21
22 (62) Kublelka, J.; Keiderling, T. A. Ab Initio Calculation of Amide Carbonyl Stretch Vi-
23 brational Frequencies in Solution with Modified Basis Sets. 1. N-Methyl Acetamide. *J.*
24 *Phys. Chem. A* **2001**, *105*, 10922–10928.
25
26
27
28 (63) Song, S. S.; Asher, S. A.; Krimm, S.; Bandekar, J. Assignment of a New Conformation-
29 Sensitive UV Resonance Raman Band in Peptides and Proteins. *J. Am. Chem. Soc.*
30 **1988**, *110*, 8547–8548.
31
32
33
34 (64) Song, S. S.; Asher, S. A.; Krimm, S.; Shaw, D. K. Ultraviolet Resonance Raman Studies
35 of Trans and Cis Peptides: Photochemical Consequences of the Twisted Π^* Excited
36 State. *J. Am. Chem. Soc.* **1991**, *113*, 1155–1163.
37
38
39
40 (65) Fang, C.; Wang, J.; Charnley, A.; Barber-Armstrong, W.; Smith, A.; Decatur, S.;
41 Hochstrasser, R. Two-Dimensional Infrared Measurements of the Coupling Between
42 Amide Modes of an Alpha-Helix. *Chem. Phys. Lett.* **2003**, *382*, 586–592.
43
44
45
46 (66) DeCamp, M. F.; DeFlores, L.; McCracken, J. M.; Tokmakoff, A.; Kwac, K.; Cho, M.
47 Amide I Vibrational Dynamics of N-Methylacetamide in Polar Solvents: The Role of
48 Electrostatic Interactions. *J. Phys. Chem. B* **2005**, *109*, 11016–11026.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (67) Li, S.; Schmidt, J. R.; Piryatinski, A.; Lawrence, C. P.; Skinner, J. L. Vibrational
4 Spectral Diffusion of Azide in Water. *J. Phys. Chem. B* **2006**, *110*, 18933–18938.
5
6
7
8 (68) Zhang, X.-X.; Jones, K. C.; Fitzpatrick, A.; Peng, C. S.; Feng, C.-J.; Baiz, C. R.;
9 Tokmakoff, A. Studying Protein-Protein Binding through T-Jump Induced Dissocia-
10 tion: Transient 2D IR Spectroscopy of Insulin Dimer. *J. Phys. Chem. B* **2016**, *120*,
11 5134–5145.
12
13
14
15
16 (69) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheese-
17 man, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian 09
18 Revision D.01. Gaussian Inc. Wallingford CT 2009.
19
20
21
22
23 (70) Colbert, D. T.; Miller, W. H. A Novel Discrete Variable Representation for Quantum
24 Mechanical Reactive Scattering Via the Smatrix Kohn Method. *J. Chem. Phys.* **1992**,
25 *96*, 1982–1991.
26
27
28
29
30 (71) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. Electric Fields at the Active Site
31 of an Enzyme : Direct Comparison of Experiment with Theory. **2006**, *313*, 200–204.
32
33
34
35 (72) Mondal, P.; Meuwly, M. Vibrational Stark Spectroscopy for Assessing Ligand-Binding
36 Strengths in a Protein. *Phys. Chem. Chem. Phys.* **2017**, *19*, 16131–16143.
37
38
39
40 (73) Lord, R. S.; Gubensek, F.; Rupley, J. A. Insulin Self-Association. Spectrum Changes
41 and Thermodynamics. *Biochemistry* **1973**, *12*, 4385–4392.
42
43
44
45 (74) Falconi, M.; Cambria, M.; Cambria, A.; Desideri, A. Structure and Stability of the
46 Insulin Dimer Investigated by Molecular Dynamics Simulation. *J. Biomol. Struct. Dyn.*
47 **2001**, *18*, 761–772.
48
49
50
51 (75) El Hage, K.; Brickel, S.; Hermelin, S.; Gaulier, G.; Schmidt, C.; Bonacina, L.; van
52 Keulen, S. C.; Bhattacharyya, S.; Chergui, M.; Hamm, P. et al. Implications of Short
53 Time Scale Dynamics on Long Time Processes. *Struct. Dyn.* **2017**, *4*, 061507.
54
55
56
57
58
59
60

Graphical TOC Entry

