Supplementary Information

Refining Protein Amide I Spectrum Simulations with Simple yet Effective Electrostatic Models for Local Wavenumbers and Dipole Derivative Magnitudes

Cesare M. Baronio, Andreas Barth*

Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

*Corresponding author email: barth@dbb.su.se

Supplementary Methods

Structures

Fig. S1 shows a selection of the structures used to optimize the parameters of the electrostatic models and Fig. S2 some auxiliary structures that either served to calculate a reference wavenumber for unperturbed amide groups (*trans N*-methylacetamide, NMA, Fig. S2a) or to assess the local conformation effect (diamides, Figs. S2b and c).
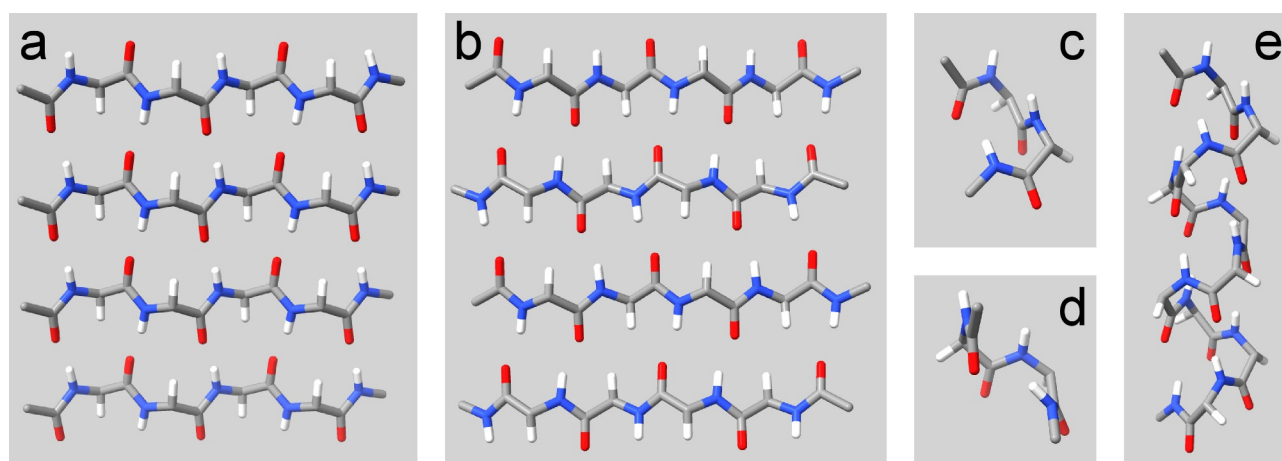

Fig. S1. Optimized structures for some of the models used to optimize the parameters of the electrostatic models. (a) 4-stranded parallel β-sheet with 5 amide groups per strand, (b) 4-stranded antiparallel β-sheet with 5 amide groups per strand, (c) and (d) two views of the α-helix with 3 amide groups of which the latter shows the non-linear hydrogen bond between amide groups 1 and 2, and (e) α-helix with 11 amide groups. The Ala side chains are omitted in all panels.
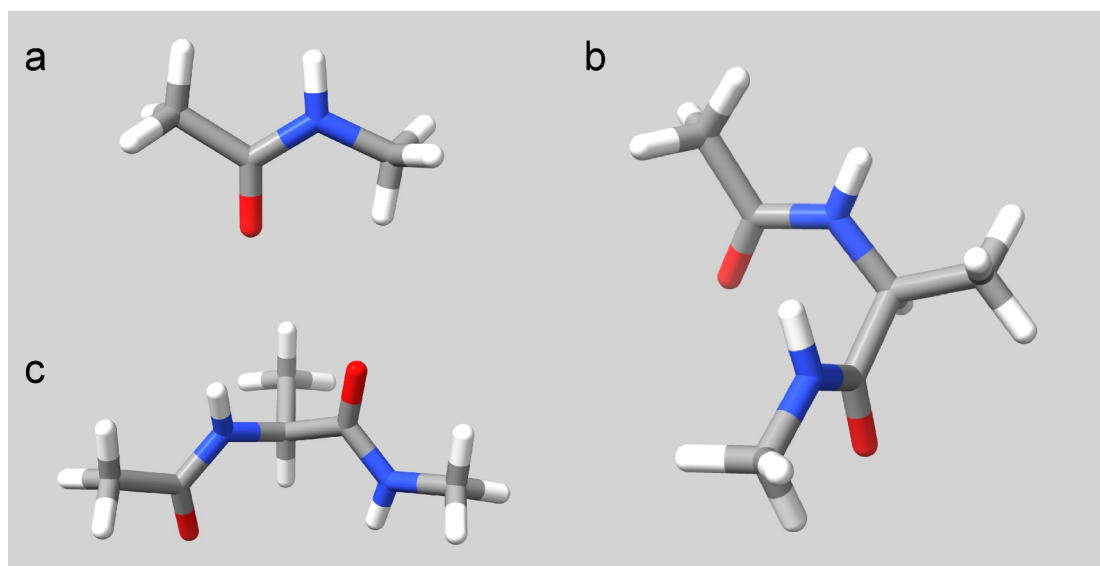
Fig. S2. Optimized structures of (a) *trans N*-methylacetamide (NMA), (b) an α-helix with two amide groups, and (c) a diamide with backbone dihedral angles as in parallel β-sheets. The respective structure for antiparallel sheets is very similar and therefore not shown.

Charge sets

Table S1 lists charge sets that performed best with a particular electrostatic model for one of the quality parameters *RMSDiff* or *RMSDev*, or served as a good compromise for both of them.

Table S1. Charge sets with good performance for particular electrostatic models.

Name of charge set[a]	Partial charge in *e* on				Net charge on CO in *e*	Net charge on CO normalized to charge on O	Relevant for electrostatic model
	C	**O**	**N**	**H**			
60/*60*/*10*/10	0.6	-0.6	-0.1	0.1	0	0.00	4P4F$_6$
Gromos: 45/*45*/*31*/31	0.45	-0.45	-0.31	0.31	0	0.00	3F$_{COH}$, 2F$_{OH}$, 2F$_{CN}$
42/*42*/*10*/10	0.42	-0.42	-0.1	0.1	0	0.00	4P
40/*40*/*10*/10	0.4	-0.4	-0.1	0.1	0	0.00	4P
50/*60*/00/10	0.5	-0.6	0	0.1	-0.1	0.17	4P4F$_8$
50/*60*/*10*/20	0.5	-0.6	-0.1	0.2	-0.1	0.17	4P4F$_6$
40/*50*/00/10	0.4	-0.5	0	0.1	-0.1	0.20	4P4F$_8$
70/*90*/00/20	0.7	-0.9	0	0.2	-0.2	0.22	4P4F$_6$
50/*65*/00/15	0.5	-0.65	0	0.15	-0.15	0.23	4F
45/*60*/00/15	0.45	-0.6	0	0.15	-0.15	0.25	3F$_{ONH}$ and LCE[b]
30/*40*/00/10	0.3	-0.4	0	0.1	-0.1	0.25	4F, 3F$_{ONH}$
Set-4P4F$_8$: 40/*55*/05/10	0.4	-0.55	0.05	0.1	-0.15	0.27	4P4F$_8$
Set-3F$_{ONH}$: 40/*55*/00/15	0.4	-0.55	0	0.15	-0.15	0.27	4F, 3F$_{ONH}$
50/*70*/00/20	0.5	-0.7	0	0.2	-0.2	0.29	3F$_{ONH}$
50/*70*/*10*/30	0.5	-0.7	-0.1	0.3	-0.2	0.29	3F$_{CON}$, 3F$_{COH}$
35/*50*/00/15	0.35	-0.5	0	0.15	-0.15	0.30	3F$_{ONH}$ and LCE[b]
40/*60*/*20*/40	0.4	-0.6	-0.2	0.4	-0.2	0.33	2F$_{ON}$, 2F$_{CN}$
35/*55*/05/15	0.35	-0.55	0.05	0.15	-0.2	0.36	4P4F$_8$ and LCE[b]
10/*70*/10/50	0.1	-0.7	0.1	0.5	-0.6	0.86	2F$_{ON}$

[a] The sets are sorted according to the net charge on the CO group normalized by the charge of the oxygen atom. A shorthand notation is used to name the charge sets and to indicate the partial charges: the partial charges in *e* are multiplied with 100 and listed in the order C/O/N/H. The minus sign for negative partial charges is omitted and instead indicated by italic type. The electrostatic models are described in the main text.
[b] LCE refers to parameter optimizations where the local conformation effect was explicitly accounted for (see main text).

Parameters of the 4P4F$_6$ model

The 4P4F$_6$ model has six free parameters although it considers the fields and potentials on all four amide atoms. The reduced number of free parameters is due to the following reasons: (i) The model relates the *l* parameters to the charge changes on each amide atom during the amide I vibration. In the absence of intermolecular charge flux, these changes should add up to zero, which implies that one of the *l* parameters can be calculated from the others. (ii) The *d* and *l* parameters are related to the dipole derivative component along the C=O bond (equation 9 in the original work[1]) and thus one of the *d* parameters can be calculated from the other parameters. We used the coordinates from our NMA calculation and Torii's *d* and *l* parameters as well as his dipole derivative component (-3.3625 D Å^{-1} a$^{-1/2}$) to calculate the quantity *b/k* in equation 9 of reference 1. A second way to calculate *b/k* used $k = 1.774$ mdyn Å^{-1} u^{-1} from our NMA calculation and calculated *b* from its definition as coefficient between wavenumber change $\Delta\tilde{v}$ and the equilibrium structural perturbation due to electrostatic interactions. The latter is expressed as a change in normal coordinate ΔQ,[1] which can be related to the change of the equilibrium C=O bond length r_{CO}: $b = (\partial\tilde{v} / \partial r_{CO})(\partial r_{CO} / \partial Q)$. The partial derivatives in this equation are given in reference 2. Both *b/k* calculations gave a value near 960 Å u$^{1/2}$ cm^{-1} E$_h$$^{-1}$, which was used in our implementation of the 4P4F$_6$ model.

Supplementary Results and Discussion

Performance of the 4P4F$_6$ and 3F$_{ONH}$ models for the inner amides

The performance of the 4P4F$_6$ model with the original parameters[1] after optimizing the partial charges on the amide atoms is listed in Table S2. The model performs considerably worse than the other models, but it turned out that the large *RMSDev* value is mainly due to the amide groups at the N- or the C-terminus of a strand or a helix. Therefore, we repeated the optimization with only those amide groups that have both an N- and a C-terminal neighbor – the inner amides. This improved the *RMSDiff* value of the inner amides considerably from 11.2 cm^{-1} to 7.0 cm^{-1} but the *RMSDev* value to a much lesser extent from 5.9 to 5.5 cm^{-1}. These values are still considerably worse that those of the 3F$_{ONH}$ model as obtained in our standard optimization that considers all amide groups. Also for the 3F$_{ONH}$ model and all other models, *RMSDiff* and *RMSDev* for the inner amide groups were smaller than for all amide groups for nearly all charge sets. This indicates that the terminal amide groups are difficult to describe with the same model and charge set as used for the inner groups. Underlying causes for this observation may be that compensating errors in the description of the local conformation effect lead to a better modeling of the inner amides than of the terminal amides or that the missing neighbor of a terminal amide generates a local wavenumber shift that requires a different description. For the 3F$_{ONH}$ model, the best charge sets for *RMSDev* of the inner groups had a higher relative net charge on the NH group than the best charge sets for all groups (NH net charge relative to oxygen charge: 50-70% for best *RMSDev* of inner groups, 30% for best *RMSDev* of all groups).

Table S2. Performance of the investigated electrostatic models.

Model	Method[a]	Number of charge sets[b]	Best charge set for	Name of charge set[c]	RMSDiff (cm^{-1})	RMSDev (cm^{-1})
4P4F$_8$	1	41	*RMSDev*	40/55/05/10 (Set-4P4F$_8$)	4.39	3.57
			RMSDiff	50/60/00/10	4.35	3.68
	2	6	*RMSDev*	40/55/05/10 (Set-4P4F$_8$)	4.33	3.40
			RMSDiff	50/60/00/10	4.28	3.45
4P4F$_8$ and LCE[d]	1	8	*RMSDev and RMSDiff*	35/55/05/15	2.44	2.19
4P4F$_6$	1	41	*RMSDev*	70/90/00/20	5.07	3.87
			Compromise	50/60/10/20	4.80	3.97
			RMSDiff	60/60/10/10	4.62	4.19
		opt. all amides[f]	*Original parameters*[g]	*19/04/16/31*	12.55, *11.18*[e]	9.84, *5.91*[e]
		opt. inner amides[f]	*Original parameters*[g]	*43/33/50/40*	15.62, *6.95*[e]	15.18, *5.54*[e]
	2	opt. all amides[f]	*Original parameters*[g]	*21/04/11/28*	12.68, *11.81*[e]	9.34, *6.02*[e]
		opt. inner amides[f]	*Original parameters*[g]	*41/32/47/38*	15.09, *6.68*[e]	14.73, *5.60*[e]
4P	1	41	*RMSDev*	40/40/10/10	5.23	4.50
			RMSDiff	42/42/10/10	5.23	4.50
4F	1	41	*RMSDev*	40/55/00/15 (Set-3F$_{ONH}$)	5.18	3.84
			Compromise	30/40/00/10	5.14	3.85
			RMSDiff	50/65/00/15	5.13	3.87
	2	6	*RMSDev*	30/40/00/10	5.06	3.82
			RMSDiff	50/65/00/15	5.04	3.83
3F$_{ONH}$	1	36	*RMSDev*	40/55/00/15 (Set-3F$_{ONH}$)	5.27, *4.61*[e]	3.88, *3.26*[e]
			RMSDiff	50/70/00/20	5.27, *4.59*[e]	3.88, *3.24*[e]
			Compromise for inner amides	50/70/10/30	5.44, *4.60*[e]	3.96, *3.15*[e]
	2	10	*RMSDev*	30/40/00/10	5.19, *4.62*[e]	3.84, *3.25*[e]
			RMSDiff	40/55/00/15 (Set-3F$_{ONH}$)	5.17, *4.56*[e]	3.85, *3.23*[e]
3F$_{ONH}$ and LCE[d]	1	9	*RMSDev*	35/50/00/15	4.68	4.31
			Compromise	40/55/00/15 (Set-3F$_{ONH}$)	4.68	4.31
			RMSDiff	45/60/00/15	4.67	4.32
3F$_{CON}$	1	5	*RMSDev* and RM*SDiff*	50/70/10/30	6.87	5.33
3F$_{COH}$	1	5	*RMSDev*	50/70/10/30	8.57	6.19
			RMSDiff	45/45/31/31 (Gromos)	8.31	7.05
2F$_{ON}$	1	36	*RMSDev*	10/70/10/50	7.46	5.00
			RMSDiff	40/60/20/40	6.79	5.20
2F$_{OH}$	1	5	*RMSDev and RMSDiff*	45/45/31/31 (Gromos)	9.08	7.09
2F$_{CN}$	1	5	*RMSDev*	40/60/20/40	10.61	8.70
			RMSDiff	45/45/31/31 (Gromos)	9.91	9.56
Zero[h]					38.5, *40.7*[e]	12.1, *10.2*[e]

[a] Column *Method* refers to the optimization method.

[b] For each model, the number of studied charge sets is listed

[c] The charge set that gave the smallest *RMSDiff* and *RMSDev* is indicated with the notation specified in Table S1. In some cases a good compromise is also listed. The best charge sets of the most relevant models were given names.

[d] LCE refers to parameter optimizations where the local conformation effect was explicitly accounted for (see main text).

[e] For some models, two values are stated for *RMSDiff* and *RMSDev*: the first value relates to all amide groups, whereas the second in italics relates only to the inner amide groups, i.e. those that have an N-terminal and C-terminal neighbor.

[f] "opt." indicates a charge set that was optimized either for all or for the inner amide groups.

[g] The original parameters for the 4P4F$_6$ model were taken from reference 1.

[h] The model named *Zero* refers to a calculation in which no wavenumber shift was applied, *i.e.* all amides had the NMA wavenumber and were compared to the DFT wavenumbers.

Charge preferences for particular electrostatic models

This section discusses charge preferences of particular models in addition to the general trends discussed in the main text. For the $4P4F_8$ model, a small positive charge on the N-atom is beneficial, whereas a negative charge is unfavorable. The best $4P4F_6$ charges have similar charge ratios as those of the $4P4F_8$ model, but the nitrogen charge is zero. The 4P model has an exceptional charge set with neutral CO and NH groups. Nevertheless, the DSSP and Gromos charge sets do not perform well because of the relatively large partial charges on the N and H atoms. In the best charge set, the hydrogen (nitrogen) charge is only 25% of the charge on the carbon (oxygen) atom, whereas it is ~70% in the Gromos charge set. The best charges for the 4F model are similar to those of the $4P4F_8$ model with the difference that the 4F model performs best when the positive charge on the NH group is entirely localized on the hydrogen atom. The performance of charge sets with the $3F_{ONH}$ model is similar to that with the 4F model and the best charge set is the same. The $2F_{ON}$ model "likes" a considerable net charge on the CO and the NH groups, with much of the charge being localized on the O and H atoms. However, concentrating all charges on these two atoms reduces the performance. The best charge set has negative charge only on the O atom, large positive partial charge on the H atom and the remaining positive charge equally distributed over the C and the N atom. Compared to the more comprehensive models, the 2F models prefer relatively larger partial charges on the N and H atoms.

Parameters of the electrostatic models

Table S3 list the parameters for the most relevant models, the $4P4F_8$, 4P, 4F and the $3F_{ONH}$ model. In the 4F model, the C-atom has least impact on the wavenumber shift and therefore it is not surprising that its omission in the $3F_{ONH}$ model left the quality measures essentially unchanged. Omitting instead the atom with second least impact, the H-atom, generated the $3F_{CON}$ model, which performed clearly worse than the $3F_{ONH}$ model. The largest impact in the 4F model had the N-atom and its omission in $3F_{COH}$ produced the worst performing 3F model that we tested. Because the performance of the 3F models followed the impact of the atoms in the 4F model, we did not test a 3F model without the O-atom, which is the second most important atom in the 4F model. Similarly, we refrained from testing the $2F_{CH}$ model, because of the bad performance of 3F and 2F models containing the C atom.

In order to explain the relative irrelevance of the C atom, we analyzed electric fields and potentials at all amide atoms in the large helix and antiparallel β-sheet structures. To highlight the effects of hydrogen bonding, we excluded the nearest sequence neighbors in the electrostatic calculations of a particular amide group. The potentials on all amide atoms and the fields at the O, N and H atoms reflected the hydrogen bonding pattern because the values were different for hydrogen bonding to only the NH group, to only the CO group, and to both groups. In contrast, the field at the C atom was less characteristic and had similar values for hydrogen bonding to only the NH group and to only the CO group. This explains the minimal loss in performance when the 4F model is replaced by the F_{ONH} model, where the carbon atom has been omitted.

Before discussing the sign and the variation of the parameters in the different models, we will first analyze a typical hydrogen bonding situation in proteins, where both the amide oxygen and hydrogen are hydrogen bonded. In this situation we have positive external charge close to the oxygen and negative charge close to the hydrogen. The potential is therefore positive close to the

oxygen and negative close to the hydrogen, implying that negative l_O and l_C parameters as well as positive l_N and l_H parameters lead to the expected downshift of the local wavenumber. However, the distance to the positive and the negative charge is more similar for the C- and the N-atom than for the O- and the H-atom, which partially cancels the contributions of the two external charges to the potential at the C- and N-atoms and makes their potential harder to predict when the structural context is more complex. In the discussed hydrogen bonding situation, the electric field across the amide group points approximately from the oxygen to the hydrogen. Its component parallel to the C=O bond is therefore negative according to our sign convention (see Methods) and positive d parameters lead to the expected downshift of the local wavenumber. In summary, a typical hydrogen bonding pattern can be modeled with positive d and l_H (and l_N) parameters and negative l_O (and l_C).

In line with these expectations, d_O and d_N are positive for all charge sets tested for the models $2F_{ON}$, $3F_{ONH}$, and 4F, as well as for the best 22 and 15 charge sets for the $4P4F_6$ and $4P4F_8$ models respectively. d_H is also positive in models where the H-atom is the only atom of the NH group that is considered ($3F_{COH}$, $2F_{OH}$) but it is negative for all charge sets tested for $3F_{ONH}$, 4F, $4P4F_6$, and $4P4F_8$. d_C is positive for the better charge sets for 4F and $4P4F_6$, but negative in particular for sets with strongly negative N atom in case of the 4F model and for sets with small partial charges in case of the $4P4F_6$ model. In contrast to the 4F model, d_C is negative for nearly all charge sets of $4P4F_8$ including the best performing half of the charge sets.

The l_O and l_H parameters are negative and positive, respectively, in all charge sets of the 4P model according to the expectation expressed above. l_H is relatively small compared to the other l parameters, which might rather reflect the strong charge on the hydrogen bond accepting oxygen than the unimportance of the potential at the H-atom. l_C and l_N have the opposite sign compared to l_O and l_H respectively. Therefore potential differences, *i.e.* the electric field, along the C=O and the N-H bond are inherent in the 4P model. The $4P4F_8$ model retains the signs of the 4P model for all l parameters with nearly all charge sets (including the best 12 sets) but puts more emphasis on the importance of the H-atom. In contrast, the $4P4F_6$ model retains only the signs for l_N and l_H, but has varying signs for l_C and l_O, with the best three charge sets having the opposite sign for these parameters than for the 4P model. Note that the l parameters in the $4P4F_8$ and 4P models were independent from each other, while previous work constrained their sum to be zero.[3] For both models, our l parameter sum was negative for all charge sets, the average sum was close to -100 e E_h^{-1} cm^{-1} (corresponding to 3-7% of the largest absolute l parameter value for the better charge sets) and there was a tendency that the deviation from zero increased for the worse performing charge sets.

It is reassuring that the parameters for the simpler electrostatic models (with up to four parameters) are consistent for different charge sets. The signs of all l parameters and of the d parameters for O, N and H are the same throughout the charge sets (with the exception of l_N that is slightly positive for one charge set). Only the sign of d_C varies and this parameter can be omitted with largely retained performance.

Models that consider the electric potential seem to reflect in part the electric field along the C=O bond – as discussed above, which is consistent with the better performance of the models that consider the electric field only. The largest model $4P4F_8$ is largely a combination of the 4F and the 4P model because most of the parameter signs are retained. Only the d_C parameter has a different sign. The signs vary most for the $4P4F_6$ model even within the best ten charge sets and the best

charge sets have the opposite sign for l_C and l_O compared to the 4P model, but there are well-performing charge sets that retain all signs of the 4F and the 4P models.

Table S3. Parameters for a selection of electrostatic models.

Model	Method[a]	Best charge set for	Charge set	d_C[b]	d_O[b]	d_N[b]	d_H[b]	l_C[b]	l_O[b]	l_N[b]	l_H[b]
4P4F$_8$	1	*RMSDev*	40/55/05/10 (Set-4P4F$_8$)	-763	1645	1072	-2119	260	-490	-1489	1620
		RMSDiff	50/60/00/10	-309	1833	421	-2315	210	-341	-1885	1927
	2	*RMSDev*	40/55/05/10 (Set-4P4F$_8$)	-371	2077	1524	-1835	-315	14	-1081	1270
		RMSDiff	50/60/00/10	356	2484	1047	-2002	-642	413	-1375	1502
4P4F$_8$ and LCE[c]	1	*RMSDev* and *RMSDiff*	35/55/05/15	-2477	-3642	-2313	414	4284	-4694	545	-132
4P4F$_6$	1	*RMSDev*	70/90/00/20	1158	1719	1744	-906	-455	560	-377	272
		Compromise	50/60/10/20	1547	1532	1128	-1434	343	97	-1327	887
		RMSDiff	60/60/10/10	1761	2285	-95	-2183	611	-47	-2361	1797
		Original parameters[d]	19/04/16/31	2941	-2663	572	548	1415	-2022	1722	-1115
4P	1	*RMSDev*	40/40/10/10					4721	-3086	-1940	221
		RMSDiff	42/42/10/10					4588	-3011	-1868	209
4F	1	*RMSDev*	40/55/00/15 (Set-3F$_{ONH}$)	384	1676	3381	-1093				
		Compromise	30/40/00/10	710	2405	4716	-1561				
		RMSDiff	50/65/00/15	534	1541	2934	-990				
	2	*RMSDev*	30/40/00/10	681	2314	4732	-1554				
		RMSDiff	50/65/00/15	515	1474	2952	-988				
3F$_{ONH}$	1	*RMSDev*	40/55/00/15 (Set-3F$_{ONH}$)		1792	3461	-1045				
		RMSDiff	50/70/00/20		1361	2691	-812				
		Compromise for inner amides	50/70/10/30		1081	2406	-702				
	2	*RMSDev*	30/40/00/10		2489	4925	-1500				
		RMSDiff	40/55/00/15 (Set-3F$_{ONH}$)		1714	3504	-1062				
3F$_{ONH}$ and LCE[c]	1	*RMSDev*	35/50/00/15		1997	1625	-442				
		Compromise	40/55/00/15 (Set-3F$_{ONH}$)		1920	1510	-420				
		RMSDiff	45/60/00/15		1850	1409	-400				

[a] Column *Method* refers to the optimization method.
[b] The d parameters are given in units of $a_0\, e\, E_h^{-1}\, \mathrm{cm}^{-1}$ and the l parameters in units of $e\, E_h^{-1}\, \mathrm{cm}^{-1}$.
[c] LCE refers to parameter optimizations where the local conformation effect has been accounted for (see text).
[d] The original parameters for the 4P4F$_6$ model were taken from reference [1]. They are not optimized in contrast to all other parameters, which is indicated by the gray background.

Comparison of our parameters with those of published electrostatic models

This section compares the parameters of published electrostatic models with our models as introduced and summarized in the main text.

The 4P4F$_6$ model was originally developed to describe the interactions of NMA with a few, explicit water molecules. Torii's original parameters[1] are included in Table S3 for comparison. It can be seen that the original parameters are different from our 4P4F$_6$ parameters in relative magnitude and often also in the sign. In particular, d_O is strongly negative and d_H is positive in the original parameters.

The latter is never the case for our 4F, 4P4F$_6$ and 4P4F$_8$ models, whereas the former parameter is only moderately negative for some of the worse performing charge sets for the 4P4F$_8$ model. The original l_N and l_H parameters have the opposite sign compared to those obtained with nearly all charge sets and our 4P, 4P4F$_6$ and 4P4F$_8$ models (the exception being a badly performing charge set for 4P with slightly positive l_N).

The models that are based exclusively on the electric potential consider different numbers of atoms. Three of them[4–6] can be compared directly to our 4P model. The other models consider either more [3,7,8] or partly other atoms.[9] The models that can be compared directly have different signs for l_C, all except one have the same sign for l_O as our 4P model and the opposite sign than our l_N, and most of those that consider the hydrogen have also the opposite sign for l_H. The 6P model by Maekawa & Ge[5] has the same signs as our 4P model for all amide atom parameters common to our model. The largest l parameter was obtained for different atoms in different models but never for the hydrogen, which is in line with our parameters. The deviations between our parameters and the published ones cannot be explained by a sensitivity of the parameter sign to the charge set as all our 41 charge sets produced the same signs for all parameters. The deviation can neither be explained by our l parameter sum being not restricted to zero. For the best 15 charge sets, the absolute value of the parameter sum was rather small, smaller than 100 $e\,E_h^{-1}\,cm^{-1}$, which corresponds to less than 5% of the largest l value (l_C).

Turning to models that exclusively consider the electric field, the 4F model by Schmidt *et al.*[7] considers all electric field components, *i.e.* those parallel and perpendicular to the CO bond. The latter contributions are always smaller than the parallel contribution (along the CO bond) but the perpendicular component in the amide plane can amount to 36% of the parallel contribution. We have not tested perpendicular components in our models as suggested in previous work on the bases of the axial symmetry of the wavenumber shifts due to hydrogen bonding.[1,2] This view is supported by the worse performance of our 4P model compared to the best 3F model: 3F$_{ONH}$. If a perpendicular component in the amide plane improved the model performance considerably, the 4P model would perform better because the potentials on the C and N atoms reflect the electric field along the CN bond, which has a component perpendicular to the CO bond. Different to our 4F model, the parameters of Schmidt *et al.* for the parallel field component have most emphasis on the carbon atom and have opposite signs compared to ours on all other three atoms. The large positive value of d_C in this model ensures the expected downshift in a typical hydrogen bonding situation within proteins, while the other d parameters reduce the downshift that is generated by d_C alone. In our 4F model, the downshift is generated by the d parameters of carbon, oxygen, and nitrogen, while d_H generates an upshift. Again, the differences between these two 4F models cannot be explained by a sensitivity of the parameter signs on the charge set, since the sign of our d parameters for O, N, and H was the same for all 41 charge sets, whereas the d_C parameter sign was somewhat more variable.

The 4F model by Schmidt *et al.*[7] was simplified to a 2F$_{CN}$ model[10] and further developed by Wang *et al.*[11] for describing local wavenumber shifts in proteins with Gromos charges. We tested a few charge sets with such a model, but it performed worse than our best 2F model – 2F$_{ON}$ – and worse than our 3F$_{ONH}$ model. The d_C parameter is twice as large in the Wang *et al.* model than in our 2F$_{CN}$ model, but both parameters are positive and lead to a downshift in a typical hydrogen bonding situation (discussed above). In our model, d_N is relatively small but in the Wang *et al.* model[11] it is substantial and negative. In spite of these differences, a typical hydrogen bonding situation would

lead to a similar shift, because in the Wang *et al.* model half of the downshift caused by d_C is compensated by an upshift due to the d_N parameter, which makes the resulting downshift similar to that calculated with our $2F_{CN}$ model where the downshift is mainly produced by the d_C parameter. In line with Wang *et al.*, we find that the $2F_{CN}$ model performs well with Gromos charges in contrast to most other models.

Local wavenumbers of diamides

We performed diamide calculations at the same level of theory as for the larger structures. Table S4 compares the local wavenumbers of our calculations with those from the literature.

Table S4. Local wavenumbers of the two amide groups in diamides reflecting the local conformation effect.

Structure	Group	C^b	G^b	LCJ^b	This work	$C_{sc}^{b,c}$	$G_{sc}^{b,c}$	$LCJ_{sc}^{b,c}$
				Local wavenumber / cm^{-1} [a]				
Helix	N^d	1727.6	1764.3	1727.7	1733.0	1739.7	1737.3	1737.3
	C^d	1734.0	1770.4	1735.2	1745.6	1746.1	1743.4	1744.8
ABS^e	N^d	1704.4	1743.5	1707.2	1717.9	1716.3	1716.9	1716.6
	C^d	1708.2	1749.5	1713.0	1721.2	1720.1	1722.8	1722.4
PBS^f	N^d	1704.7	1743.7	1708.0	1720.6	1716.6	1717.1	1717.4
	C^d	1711.1	1751.2	1713.9	1723.5	1723.1	1724.4	1723.3
Factor						1.0070	0.9847	1.0055
RMSDev / cm^{-1} [g]						3.3	2.6	2.3

[a] The local wavenumbers were obtained from the DFT calculations by Hessian matrix reconstruction.
[b] Published data are indicated by the initials of the first author, C,[12] G,[13] LCJ.[14,15] See Methods of the main text for further details.
[c] The subscript "SC" indicates values that were scaled with a common factor for each data set so that the average wavenumber was the same as for our calculations.
[d] N and C in the Group column refer to the N-terminal and C-terminal amide groups.
[e] ABS: antiparallel β-sheet,
[f] PBS: parallel β-sheet.
[g] *RMSDev* is the root mean square deviation of the published data from our values.

Local wavenumber shifts for the 3F$_{ONH}$ + LCE model

Fig. S3 compares local wavenumber shifts of individual amide groups for the 3F$_{ONH}$ model with or without explicit consideration of the local conformation effect which are discussed in the main text.
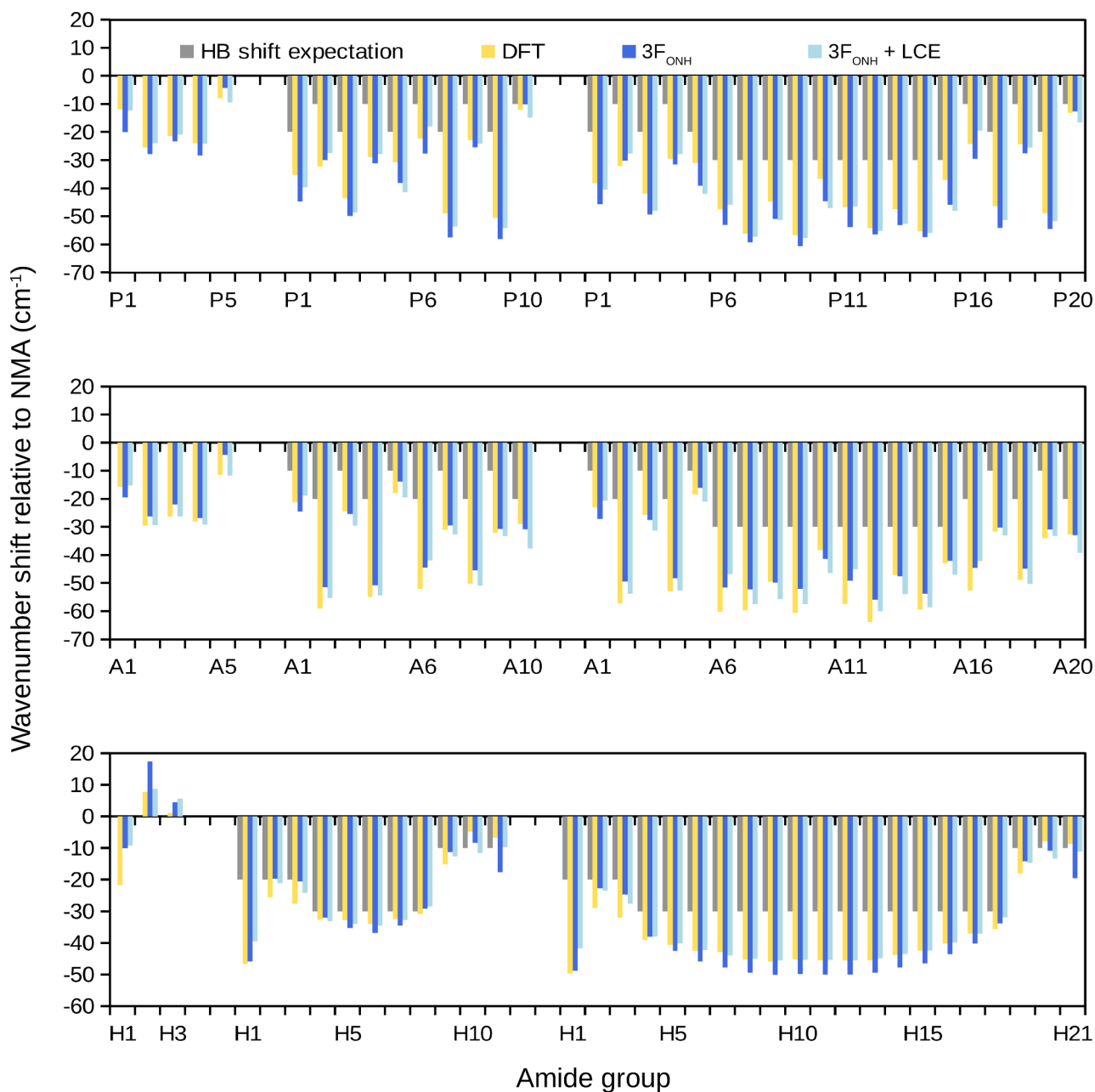

Fig. S3. Local wavenumber shifts of individual amide groups calculated by DFT (golden bars), the 3F$_{ONH}$ model (blue bars), and the 3F$_{ONH}$ + LCE model (light blue bars) using Optimization 1 (Table 1). Top: parallel β-strand and β-sheets (P), middle: antiparallel β-strand and β-sheets (A), bottom: helices (H). The horizontal axis displays the number of the amide groups, where 1 indicates the N-terminal amide group. Small, medium, and large structures are arranged from left to right. A simple indication of the hydrogen bonding type is shown by gray bars (0, -10, -20, and -30 cm^{-1} shift for hydrogen bonds to neither H nor O, to H, to O, and to H and O, respectively).

Sensitivity of the model parameters on the explicit consideration of the local conformation effect

Table S3 lists the parameters of the $4P4F_8$ and $3F_{ONH}$ models obtained by optimization with and without explicit consideration of the local conformation effect. The parameters of the $4P4F_8$ model change drastically: Half of them change sign and their absolute values are also significantly different. Strikingly, the sum of the d parameters is negative and at least 4 times larger for optimizations with local conformation effect compensation when the better performing charge sets are compared (*e.g.* Set-$4P4F_8$, Set-$3F_{ONH}$, and 35/*55*/05/15), but not for the Gromos and DSSP charge sets. The d parameter sum of $4P4F_8$ with local conformation compensation ($4P4F_8$ and LCE) has a larger absolute value than the d parameter sum of all 4F and $3F_{ONH}$ models and has the opposite sign. It generates a strong upshift in a homogeneous electric field that corresponds to a typical hydrogen bonding situation with hydrogen bonds to both the CO and the NH groups. This is in contrast to the known effects of hydrogen bonding and needs therefore to be over-compensated by the potential term and the associated l parameters. We consider such a model as risky because two large local wavenumber shifts (from the electric field and the electric potential) in opposite directions partly compensate each other to generate a smaller resulting wavenumber shift. Small errors in the shifts caused by the electric field and the electric potential will therefore generate a large error in the resulting wavenumber shift.

In contrast, the parameters of the $3F_{ONH}$ model change relatively little and all signs are retained when the local conformation effect is explicitly considered. For the better performing charge sets, the field on the O atom is approximately as important (similar d_O), but the fields on the N and the H atom are less significant (d_N and d_H are approximately halved) when the local conformation effect has been compensated for. This seems to indicate that the fields on the N and H atoms are important to describe the local conformation effect.

Performance of combining a short-range hydrogen bonding model with an explicit consideration of the local conformation effect

The main text summarizes the performance of combining a short-range hydrogen bonding model with an explicit consideration of the local conformation effect. This combined model is discussed in more detail in the following. The shifts from the local conformation effect are compiled in Table S5 for two published data sets and our diamide calculations. They were obtained as described in Methods of the main text. The hydrogen bonding model did not detect an interaction between nearest sequence neighbors, thus this interaction is entirely described by the local conformation effect. Using our set of nine structures, the shifts were optimized by applying offsets to the maps of the local conformation effect due to the preceding and the following amide group. The offsets reflect that the appropriate reference wavenumber – used to generate the shifts from the local wavenumbers in diamides – is not exactly known. Consequently, different wavenumbers have been used for the published maps (see Methods of the main text). The offsets may also reflect electrostatic effects not described by the hydrogen bonding model, for example the effect of the helix dipole. In our optimization, we used either the same optimized offset for both maps or individually optimized offsets for each map. We optimized also the coefficients ξ^O and ξ^H that link the Kabsch-Sander energy to a wavenumber shift. The performance of these models is compiled in Table S6.

There is little difference between the optimizations with a common offset for both maps and with different offsets for the two maps. This is because the sum of the two offsets is very similar (within 0.3 cm^{-1}) for each map origin (G, LCJ, this work) which results in the same offset for the inner amide groups independent of whether a common offset or two different offsets were used.

Table S5. Local wavenumber shifts due to the local conformation effect calculated as described in Methods of the main text.

Structure	Group	Shift due to the local conformation effect in cm^{-1}		
		G	LCJ	This work
Helix	N	4.4	10.7	-2.0
	C	10.3	18.2	10.6
	Sum	14.7	28.9	8.7
ABS[a]	N	-15.8	-9.8	-17.1
	C	-10.0	-4.0	-13.8
	Sum	-25.7	-13.9	-30.8
PBS[b]	N	-15.6	-9.0	-14.4
	C	-8.4	-3.1	-11.5
	Sum	-23.9	-12.2	-26.0

See the text and Tables 2 and S4 for further details. Discrepancies between the individual shift values for the N- and C-terminal amide group and their sums are due to rounding.
[a] ABS: antiparallel β-sheet,
[b] PBS: parallel β-sheet.

Compared to the electrostatic models, *RMSDiff* is much worse with the local hydrogen bonding model, whereas *RMSDev* is close to our main models 4P4F$_8$ and 3F$_{ONH}$, but still worse. The latter indicates that a short-range hydrogen bonding model captures main aspects of the local wavenumber variations within a given structure relatively well, but that a more delocalized electrostatic description works better.

Fig. S4 shows the shifts obtained for the individual amide groups with the approach that uses our diamide data, different offsets for the two maps, and Optimization 1 (indicated by superscripted "f" in Table S6). It is evident that the shifts calculated for the sheets are too large, whereas those for the helices are too small. The varying shifts due to the alternating hydrogen bonding pattern of the terminal strands of the sheets are well reproduced, but not the larger shifts for the central helix portion when the helix length increases from 11 to 21 amide groups. *RMSDev* is small even with the standard parameters, which indicates that the original parameters of the local hydrogen bonding model describe the varying hydrogen bonding pattern approximately as well as after optimization. But it does not seem to be sufficient to describe the local wavenumbers of helices, probably because of an effect of the helix dipole.

The amides of the small β-strand structures and the inner amide of the small helix are not influenced by the hydrogen bonding model and their shifts reflect therefore exclusively the local conformation. The shifts are more negative than the DFT shift in all cases (Fig. S4), which is due to the negative offsets. The offsets are smaller with Optimization method 2 (Table S6) which brings the shifts closer to the DFT values. When they are absent in the calculation with the standard parameters, the agreement is close to perfect. However, these improvements come at the expense of large *RMSDiff* values for the large helix. It seems that the offset is used to compensate for a deficiency of the local

hydrogen bonding model to describe more long-range interactions, for example the effect of the helix dipole.

Finally, we tested a model that ignores the local conformation effect and generates local wavenumber shifts only with the local hydrogen bonding model (last row in Table S6). Interestingly, this model has the best *RMSDiff* value but the worst *RMSDev* value. Also, it has the largest Kabsch Sander parameters. The high *RMSDev* value shows that these parameters badly describe the variations of the local wavenumbers within a given structure, but the relatively low *RMSDiff* value indicates that the average shift difference is relatively well described and thus linked to the strength of the hydrogen bonds. Nevertheless, the better electrostatic models outperform this simple model by far regarding both quality parameters.

Table S6. Performance of combining a short range hydrogen bonding model with the local conformation effect.

Data	Method[a]	Offset[b] in cm^{-1} for N-terminal amide	C-terminal amide	ξ^O (cm^{-1} mol/kJ)	ξ^H (cm^{-1} mol/kJ)	*RMSDiff*[c] (cm^{-1})	*RMSDev*[c] (cm^{-1})
	1	-8.79	-8.79	2.07	0.30	14.7, *16.5*	6.2, *4.3*
	2	-4.21	-4.21	2.38	0.59	14.1, *15.7*	5.0, *4.0*
G$_{opt}$[d]	1	-12.00	-5.68	2.05	0.32	14.6, *16.5*	6.0, *4.3*
	2	-7.07	-1.41	2.35	0.61	14.0, *15.7*	4.7, *4.0*
	Std[e]	0	0	2.40	1.00	14.9, *16.7*	4.7, *4.3*
	1	-15.49	-15.49	2.05	0.28	15.7, *17.7*	6.4, *4.3*
	2	-10.61	-10.61	2.37	0.60	15.0, *16.8*	5.0, *4.0*
LCJ$_{opt}$[d]	1	-18.92	-12.16	2.02	0.30	15.6, *17.7*	6.2, *4.3*
	2	-13.63	-7.67	2.34	0.61	14.9, *16.8*	4.8, *4.0*
	Std[e]	0	0	2.40	1.00	21.6, *25.0*	5.4, *4.3*
	1	-5.99	-5.99	2.12	0.36	13.6, *15.2*	6.0, *4.2*
	2	-1.96	-1.96	2.35	0.61	13.0, *14.5*	5.0, *4.0*
This work$_{opt}$[d]	1[f]	-10.45	-1.80	2.08	0.38	13.5, *15,3*	5.7, *4.2*
	2	-5.66	1.66	2.32	0.63	12.9, *14.5*	4.6, *4.0*
	Std[e,g]	0	0	2.40	1.00	13.7, *14.6*	4.9, *4.2*
No LCE[h]	1	-	-	3.16	1.47	10.8, *11.1*	7.8, *5.2*

[a] Column *Method* refers to the optimization method.

[b] The offset for the N- or C-terminal amide corresponds to a map offset of the effect of the following or preceding group, respectively.

[c] Two values are given for *RMSDiff* and *RMSDev*: the first value relates to all amide groups, whereas the second in italics relates only to the inner amide groups, *i.e.* those that have an N-terminal and C-terminal neighbor.

[d] The subscript "opt" refers to our parameter optimization that included optimized offsets for the local conformation effect. We used either the same offset or different offsets for the N- and C-terminal amide groups. For a given data set (G, LCJ, or this work) and for all secondary structures, the same map offsets were added to the local conformation shifts listed in Table S5. In the optimization, the maps together with the local hydrogen bonding model were applied to approximate the local DFT wavenumbers of the amide groups in our nine structures.

[e] *Std* refers to a calculation with non-optimized standard parameters and without map offset.

[f] results obtained with this model are shown in Figs. 2, 3, and S4.

[g] spectra obtained with this model are shown Figs. 2 and 3.

[h] *No LCE* refers to an optimization, where the local conformation effect was not considered and the local wavenumber shift was entirely produced by the local hydrogen bonding model.
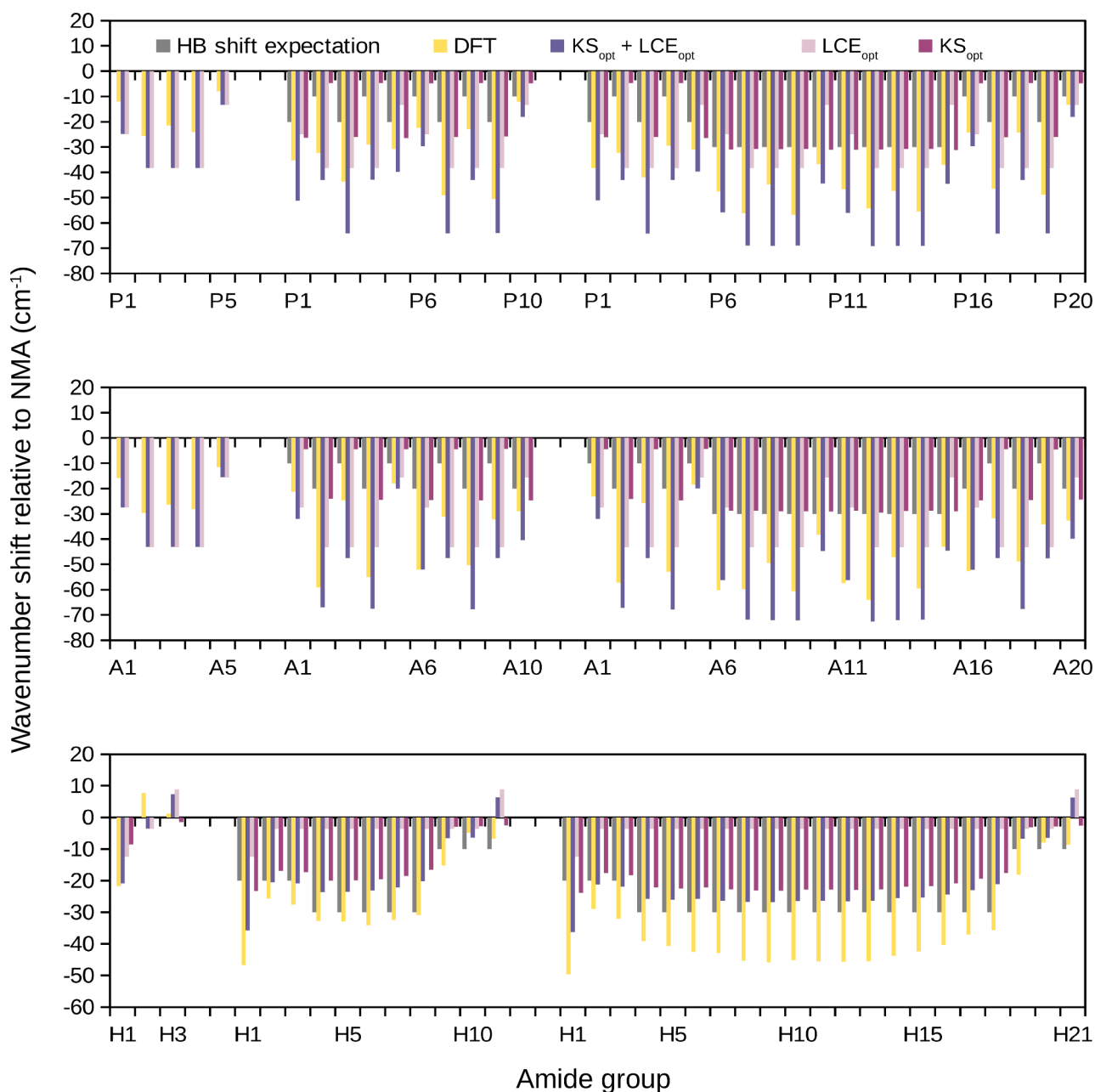
Fig. S4. Local wavenumber shifts of individual amide groups calculated by DFT (golden bars) and the combination of a local hydrogen bonding model with an explicit description of the local conformation effect (indigo bars). The local conformation effect used the diamide calculations of this work after applying an optimized map shift using Optimation 1 (Table S6). The individual contributions of the local conformation effect (light purple bars) and of the hydrogen bonding model (purple bars) are also shown. See Fig. S3 for further details. Top: parallel β-strand and β-sheets, middle: antiparallel β-strand and β-sheets, bottom: helices.

Influence of the charge set on the spectrum

Because the charge set has an important influence on the performance of the electrostatic models, we tested its influence on the spectrum. Figs. S5 and S6 compare complete spectra (from diagonal and non-diagonal *F*-matrix elements) as shown in Fig. 3 of the main text that were calculated with the same electrostatic model for different charge sets. The parameters were optimized for each charge set and the resulting quality parameters are listed in Table S7. Fig. S5 shows such results for the $3F_{ONH}$ model. The best charge set for this model is Set-$3F_{ONH}$ (red curves) but most of the other charge sets considered in Fig. S5 turn out similar quality measures *RMSDiff* and *RMSDev* and thus give very similar spectra. Clear differences are only observed for the Gromos charge set, which is the only one with clearly worse quality measures. This leads to slightly distorted spectral shapes compared to the spectra calculated from the DFT *F*-matrix. Thus, the spectrum generated by the $3F_{ONH}$ model seems to be robust against moderate changes of the amide charges including increasing the negative charge on the oxygen, varying the positive charge on the hydrogen and introducing positive or negative charges on the nitrogen.

Table S7. Charge sets for the spectra shown in Figs. S5 and S6 and the quality parameters *RMSDiff* and *RMSDev* for the $3F_{ONH}$ and $4P4F_8$ models with parameters optimized by method 1.

	$3F_{ONH}$		$4P4F_8$	
Charge set[a]	***RMSDiff*** **(cm^{-1})[b]**	***RMSDev*** **(cm^{-1})[b]**	***RMSDiff*** **(cm^{-1})[b]**	***RMSDev*** **(cm^{-1})[b]**
40/*55*/00/15 (Set-$3F_{ONH}$)	**5.27**	**3.88**	4.41	3.68
40/*60*/00/20	5.36	3.91	4.47	3.72
40/*55*/05/10 (Set-$4P4F_8$)	5.31	3.97	**4.39**	**3.57**
40/*60*/*10*/30	5.62	4.03	4.55	3.90
45/*45*/*31*/31 (Gromos)	6.53	4.90	4.96	4.16

[a] Charge sets are listed according to the *RMSDev* value of the $3F_{ONH}$ model.
[b] The best quality parameter values for each electrostatic model are highlighted by bold print.

Fig. S6 shows respective spectra for the $4P4F_8$ model showing again that the spectral shape is only little influenced by small changes of the charges. The Gromos set performs better (see the medium APS and small helix results) with the $4P4F_8$ model than with the $3F_{ONH}$ model, both regarding the spectra and the quality measures, likely because of the larger number of adjustable parameters which compensate for the deficiencies of this charge set.
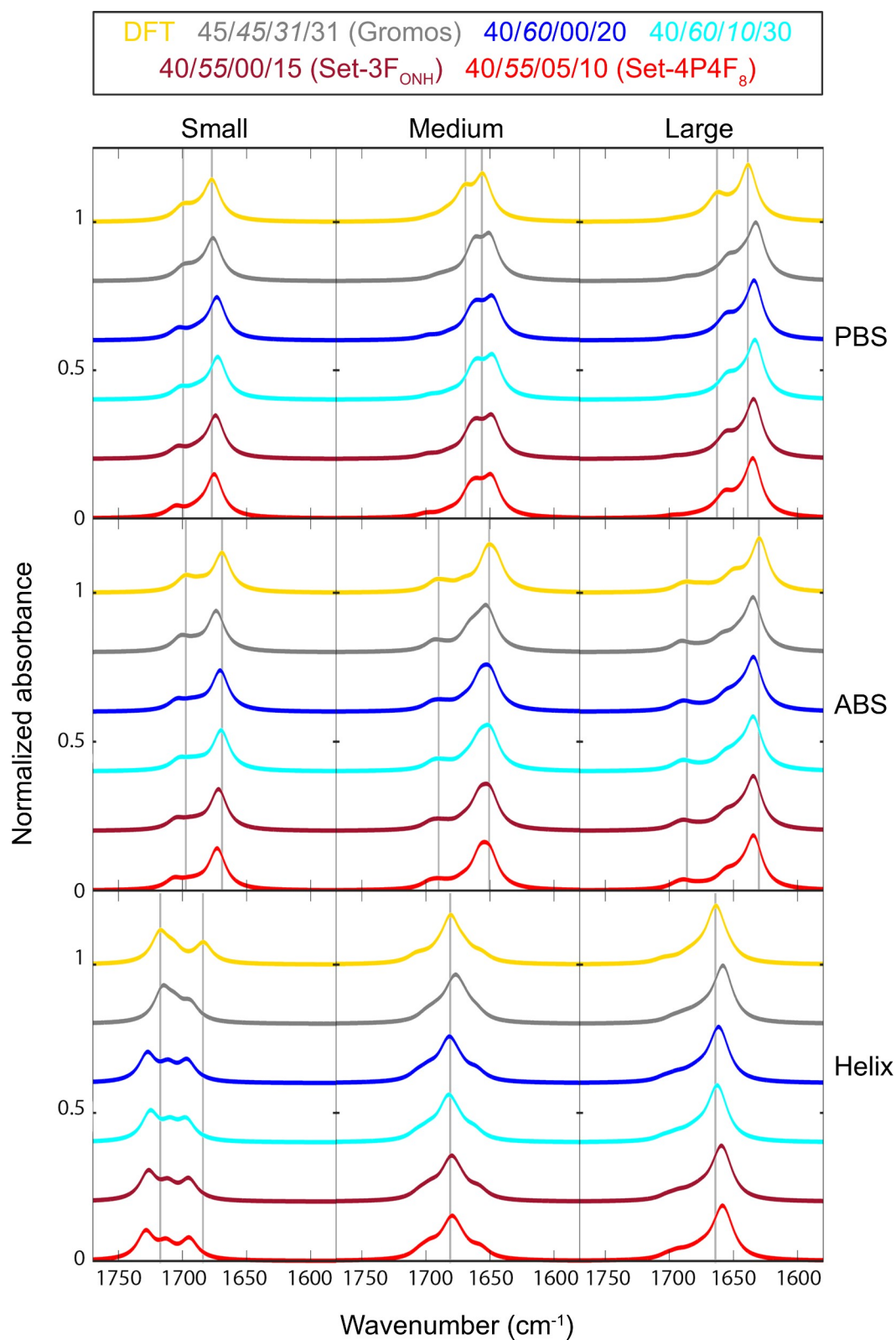
Fig. S5. Charge set influence on spectra calculated with the 3F$_{ONH}$ model. Diagonal elements of the *F*-matrix were either taken from the DFT *F*-matrix, or calculated according to the 3F$_{ONH}$ model optimized for each charge set. The non-diagonal elements were always from the DFT *F*-matrix. Top: parallel β-sheets, middle: antiparallel β-sheets, bottom: helices. The charge sets were 45/*45*/*31*/31 (Gromos, gray), 40/*60*/00/20 (blue), 40/*60*/*10*/30 (light blue), 40/*55*/00/15 (Set-3F$_{ONH,}$ dark red), and 40/*55*/05/10 (Set-4P4F$_8$, red). The gray vertical lines indicate band positions in the spectrum obtained with the DFT *F*-matrix.
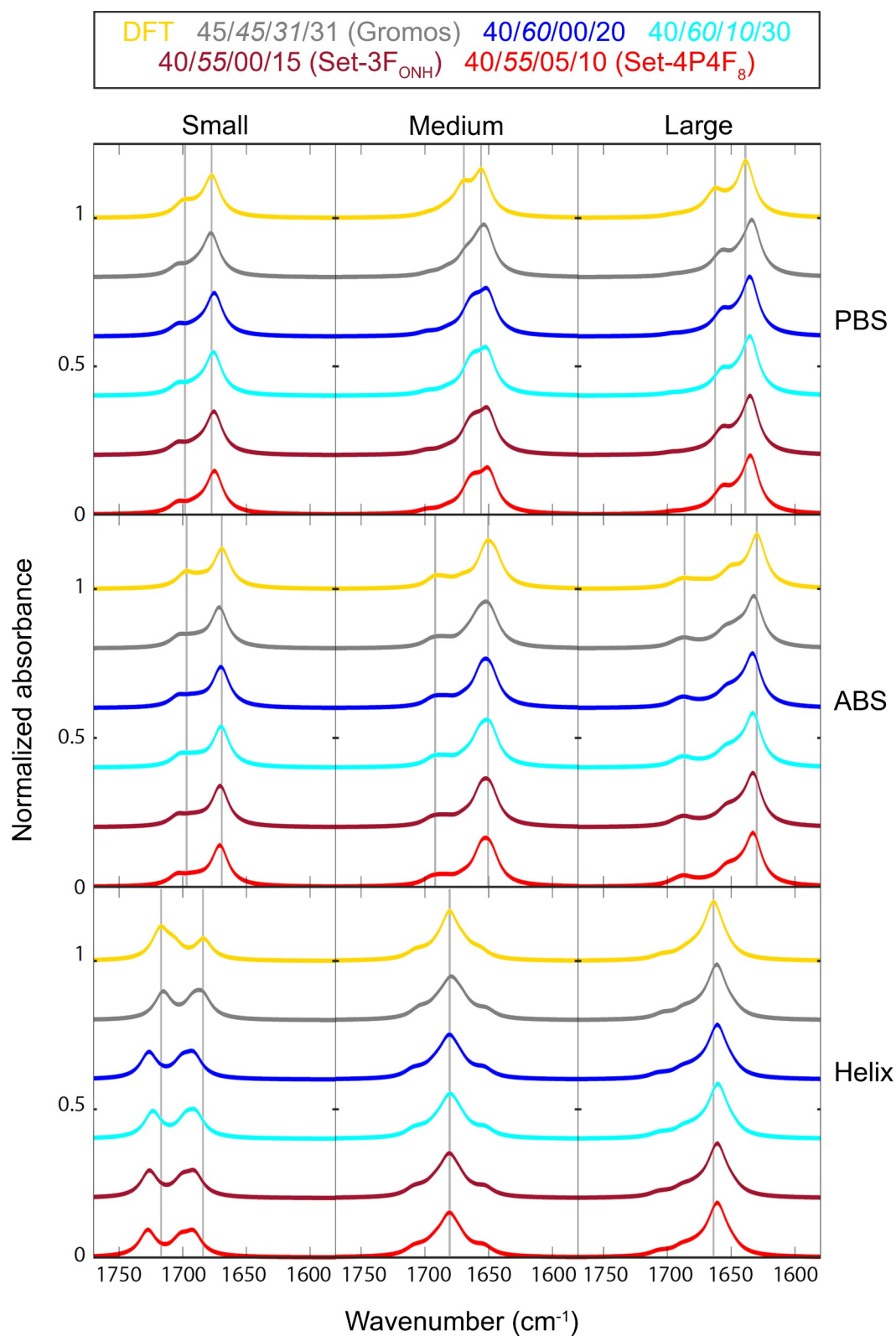
Fig. S6. Charge set influence on spectra calculated with the 4P4F$_8$ model. Diagonal elements of the F-matrix were either taken from the DFT F-matrix, or calculated according to the 4P4F$_8$ model optimized for each charge set. The non-diagonal elements were always from the DFT F-matrix. Top: parallel β-sheets, middle: antiparallel β-sheets, bottom: helices. The charge sets were the same as in Fig. S5. The gray vertical lines indicate band positions in the spectrum obtained with the DFT F-matrix.

Acknowledgments

References

1   H. Torii, *J. Phys. Chem. B*, 2018, **122**, 154–164.

2   H. Torii, *J. Phys. Chem. Lett.*, 2015, **6**, 727–733.

3   S. Ham, J.-H. Kim, H. Lee and M. Cho, *J. Chem. Phys.*, 2003, **118**, 3491–3498.

4   S. Ham, S. Cha, J.-H. J.-H. Choi and M. Cho, *J. Chem. Phys.*, 2003, **119**, 1451–1461.

5   H. Maekawa and N. Ge, *J. Phys. Chem. B*, 2010, **114**, 1434–1446.

6   T. M. Watson and J. D. Hirst, *Mol. Phys.*, 2005, **103**, 1531–1546.

7   J. R. Schmidt, S. A. Corcelli and J. L. Skinner, *J. Chem. Phys.*, 2004, **121**, 8887–8896.

8   P. Bouř, D. Michalík and J. Kapitán, *J. Chem. Phys.*, 2005, **122**, 144501.

9   P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2003, **119**, 11253–11262.

10   Y.-S. Lin, J. M. Shorb, P. Mukherjee, M. T. Zanni and J. L. Skinner, *J. Phys. Chem. B*, 2009, **113**, 592–602.

11   L. Wang, C. T. Middleton, M. T. Zanni and J. L. Skinner, *J. Phys. Chem. B*, 2011, **115**, 3713–3724.

12   J.-H. Choi and M. Cho, *J. Chem. Phys.*, 2004, **120**, 4383–92.

13   R. D. Gorbunov, D. S. Kosov and G. Stock, *J. Chem. Phys.*, 2005, **122**, 224904.

14   T. la Cour Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst and J. Knoester, *J. Chem. Phys.*, 2006, **125**, 044312.

15   T. la Cour Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst and J. Knoester, *J. Chem. Phys.*, 2012, **136**, 209901.

16   E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin, *Protein Sci.*, 2021, **30**, 70–82.