ORIGINAL PAPER

The importance of secondary structure in determining CO$_2$-protein binding patterns

Michael L. Drummond · Angela K. Wilson · Thomas R. Cundari

Abstract One potential means to decrease the level of atmospheric carbon dioxide is through the utilization of protein–CO$_2$ interactions. A recent bioinformatics analysis [Cundari TR et al. (2009) J Chem Inf Model 49:2111–2115] of these interactions revealed a marked disparity in CO$_2$ affinity between α-helices and β-sheets. In order to understand this difference, a series of molecular dynamics simulations was performed on polypeptide model systems. Numerous factors that may influence CO$_2$ affinity were systematically investigated, including the specific location of the amino acids within the secondary structural elements (SSEs), the partial charges on CO$_2$, chemical modifications made to the protein backbone, the inclusion of singly, doubly, and many functionalized residues, and the effect of solvent water. The differing abilities of the secondary structure types to participate in hydrogen bonding along the backbone were identified as being a crucial influence on CO$_2$ affinity; the lesser role of polypeptide–CO$_2$ electrostatic interactions was also noted. The effect of incorporating functionalized amino acid side chains, such as those possessed by Arg and His, on the affinity differs between the two structure types, and also strongly depends on the number included and the distance between them. The inclusion of explicit water molecules was found to attenuate all interactions, but did not change the overall trends in CO$_2$ affinity. Collectively, these results highlight the role of the backbone atoms in binding the CO$_2$ ligand, which will have important implications for efforts to ameliorate atmospheric carbon dioxide through the use of natural, designed, and modified proteins.

Introduction

Hydrogen bonds between the backbone atoms of a protein chain are responsible for determining the secondary structure of a protein, and thus to a large extent determine its function. Indeed, approximately 70% of the hydrogen bonds found in a protein are between backbone residues [1]. These intramolecular interactions are so important for maintaining the integrity of the native structure of a protein that it is exceedingly rare to find even a single hydrogen-bond donor or acceptor in a protein that is unfulfilled by water (or a ligand, e.g., CO$_2$) or the protein itself [2]. Recently, it has been suggested that the dominant force determining the pathway of protein folding is not the wide variety of interactions possible among the 20 chemically diverse amino acid side chains, but rather the strength of the network of backbone–backbone hydrogen bonds that are established as an unfolded protein spontaneously assumes an native ordered state [3].

In addition to their crucial role in determining the structure and function of a protein, backbone interactions also play a key role in mediating both protein–protein and protein–ligand contacts. The importance of the backbone in protein–protein interactions has been demonstrated both experimentally [4] and computationally [5], where it was

M. L. Drummond (✉) · A. K. Wilson · T. R. Cundari
Department of Chemistry, Center for Advanced Scientific Computing and Modeling (CASCaM), University of North Texas, 1155 Union Circle, #305070, Denton, TX 76203-5017, USA
e-mail: Michael.drummond@unt.edu

shown that the average energetic contribution of the backbone to binding is nearly identical to the contribution of the side chains [5]. Perhaps the definitive example of the power of intermolecular backbone interactions is in the formation of amyloid fibrils [6, 7], agents that are strongly implicated in neurodegenerative disorders such as Alzheimer's, Parkinson's, and Huntington's diseases [8]. The key initial feature common to these afflictions is a conformational change (a misfolding) of a protein segment [9], which can lead to the exposure of the edges of β-sheets. These newly revealed edges can form intermolecular protein–protein aggregates, analogous to the intramolecular interactions found in native β-sheets. Nature has evolved numerous structural motifs to inhibit the agglutinative capability of β-sheet edges, such as β-barrels, β-helices, β-bulges, and the Greek key, among others [10]. Scientists have also developed synthetic modifications to ameliorate β-sheet edge–edge interactions, such as esterification [11] (replacement of backbone –NH– with –O–) and methylation [12] (replacement of backbone amide –H with –CH$_3$).

A protein's backbone functionality can also be utilized to bind ligands. A striking example can be seen in the case of the chemical denaturant urea [13–15], where protein unfolding occurs specifically because the protein backbone atoms interact more strongly with this ligand than they do with other backbone atoms (or with water). The backbone–backbone hydrogen-bonded network responsible for keeping a protein in its folded state therefore unravels to take advantage of this new, more energetically favorable, interaction.

It has previously been noted [16, 17] that the specific arrangement of the backbone atoms—i.e., the secondary structural elements (SSEs)—can influence this ability of the backbone to interact with ligands. In particular, it was observed that, due to their tightly coiled structure, α-helices at catalytic sites do not possess hydrogen-bond donors capable of interacting with ligands, in contrast to the edges of β-sheets. This finding will be explored in this work in the context of protein–CO_2 interactions, an interaction of great interest in light of concerns about anthropogenic global warming. Indeed, the most abundant enzyme on the planet—ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO)—utilizes CO_2 as its carbon source for the production of glucose via the Calvin cycle [18]. In addition to serving as the ultimate foundation for most life, the protein–CO_2 interaction exhibited by RuBisCO is also responsible for removing 10^{11} tons of atmospheric carbon dioxide annually. Efforts to bioengineer an increase in the enzymatic efficiency of RuBisCO to combat rising atmospheric CO_2 concentrations have met with only moderate success [19], and thus other protein–CO_2 systems have been investigated for use in carbon capture and sequestration (CCS). These efforts [20, 21] have largely focused on

carbonic anhydrase (CA), an enzyme which captures CO_2 as aqueous bicarbonate with textbook high efficiency [22]. However, this anion is typically sequestered as $CaCO_3$, therefore requiring both a calcium counterion source—which, if cheap and abundant seawater is used, demands tremendous water flow [20]—and a means to dispose of the calcium carbonate mineral product, which is heavy and already (over)abundant. Thus, the identification of a protein framework utilizing CO_2 that is both more efficient than RuBisCO and generates a more useful value-added product than does CA would be invaluable for bioinspired or biomimetic CCS.

Toward that end, we have previously examined [23] the protein–CO_2 interaction patterns found among about 20 protein–CO_2 structures that have been characterized with X-ray crystallography. Our bioinformatics-driven approach revealed that these varied proteins—primarily (but not exclusively) enzymes, and spanning four of the six enzyme classes [24]—interact with CO_2 largely via basic amino acid residues (i.e., Arg, His, and Lys) and preferentially via β-sheets rather than α-helices. In our original work, we investigated three possible causes of this unexpected disparity in CO_2–SSE binding preference: the relative populations of the two SSEs, which actually favor α over β by a 3:2 ratio; the propensities [25] for the most CO_2-philic amino acids (Arg, His, and Lys [23]) to be disproportionately found in β-sheets rather than α-helices, which was not the case; and the rigidities of the two structural elements, which were found to be essentially identical. A final possible explanation for the β-over-α disparity is the differing abilities of the protein backbones of these two SSEs to interact with the CO_2 ligand, as discussed above (and elsewhere [16, 17]).

In this work, this hypothesis was explicitly tested via extensive molecular dynamics (MD) simulations of model α-helix and (antiparallel) β-sheet polypeptides (Fig. 1). In order to isolate the influence of the backbone structure (i.e., to remove the effect of side-chain functionality), control MD simulations were performed using all-Leu homopolypeptides. Leucine was chosen both because it can be incorporated into both α and β SSEs with equal facility [25], and because it is not commonly found at crystallographically determined CO_2 binding sites [23], and thus serves as an ideal blank. However, amino acid side-chain functionality obviously influences CO_2 affinity, and thus the two most CO_2-philic amino acids, Arg and His [23], were grafted onto the homo-Leu polypeptides in a systematic fashion in subsequent MD simulations. Thus, despite the fact that these model simulations only incorporate three amino acid types, they cover the extremes of CO_2 binding affinity. In addition, numerous other factors that may influence backbone–CO_2 interactions were probed, including the partial charges on CO_2, synthetic backbone
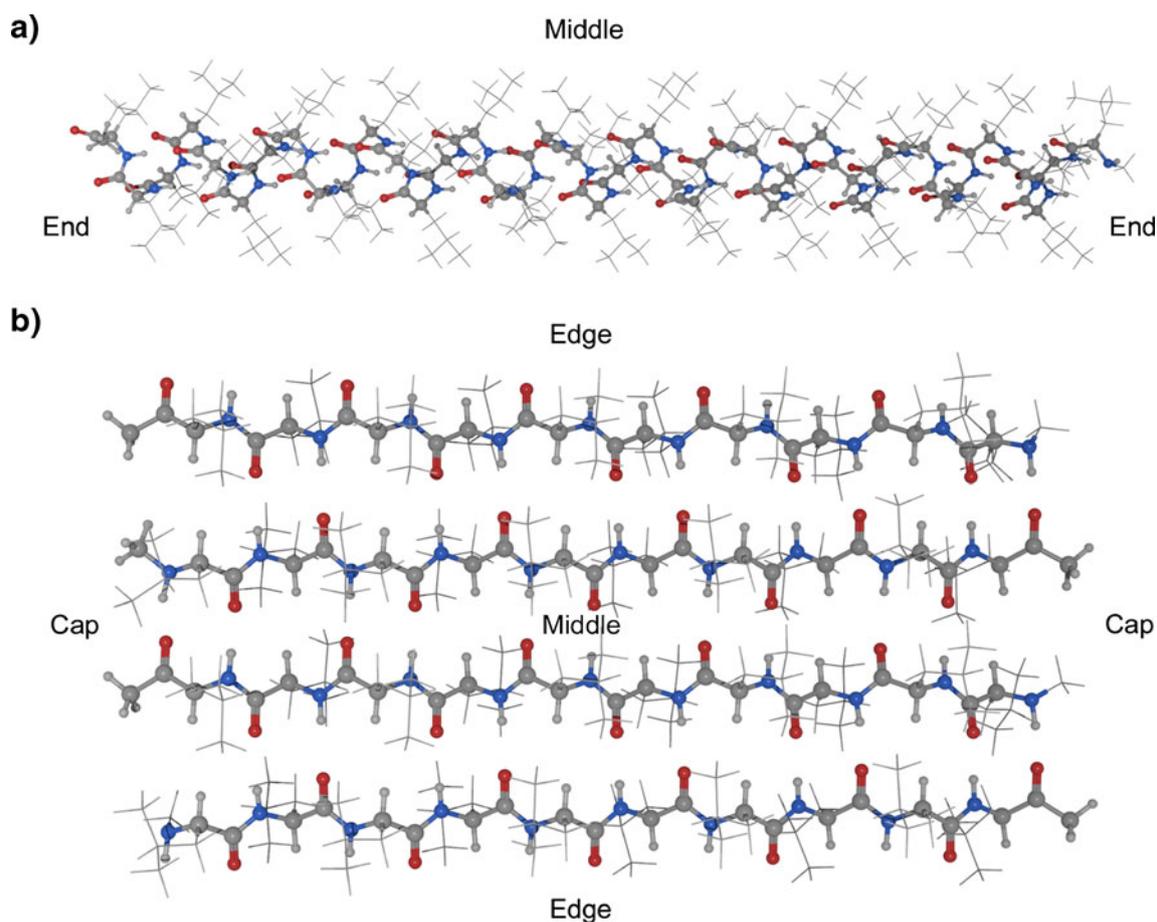
Fig. 1 The two model polypeptides used as the basis for the MD simulations in this work. Each consists of 40 residues (generally Leu, whose side chains are indicated above in "line" representation), with the backbone atoms (shown in "ball-and-stick" representation) held fixed unless otherwise noted. **a** The α-helix. *End* refers to the two N- and C-terminal residues, while the remaining 38 residues are collectively labeled *Middle*. **b** The antiparallel β-sheet. *Edge* refers to the 20 residues in the outermost two strands, *Cap* refers to the outermost four residues on the two inner strands, and *Middle* refers to the remaining 16 residues. In order to model the incorporation of these secondary structures into a larger protein, the end and cap residues were capped with methyl groups, as were the four edge residues located on strand ends

modifications such as esterification [11] or methylation [12], the identity, number, and separation of the functionalized side-chain groups, and the presence of water. Water, given its ubiquity in both protein environments and flue stacks, as well as its comparable size, is the natural choice to investigate protein–ligand interactions that might compete with CO_2.

All in all, this fundamental study provides a comprehensive investigation of the influence of SSEs on CO_2 affinity, thereby demonstrating that the disparity noted in our previous bioinformatics study [23] is not happenstance, but rather a reflection of the fact that β-sheets possess backbone hydrogen-bonding abilities that α-helices do not. Nevertheless, it is naïve to believe that isolated β-sheets will be able to effectively capture CO_2 by themselves; this is not our contention. However, one important implication of this study is that enzymes with a preponderance of β-sheets should be more effective at trapping CO_2 vis-à-vis similar enzymes with a larger percentage of α-helices. In other words, if an enzyme (e.g., carbonic anhydrase [20, 21]) is to be pursued as a means to enzymatically transform gaseous CO_2 into a more environmentally benign (and ideally useful) species, efforts should focus on the isoform with a greater proportion of β-sheets, or artificial techniques should be applied to engineer additional β-sheet functionality. In this light, the findings of this study will generally have repercussions for research geared toward selecting [24], modifying, or designing proteins with enhanced CO_2 affinity for use as "green" carbon capture tools in the fight against anthropogenic global warming.

Computational methods

All MD simulations were performed using the Molecular Operating Environment (MOE) software package, v.2008.10 [26]. The AMBER94 force field [27] was used

for all simulations; any missing parameters were automatically generated by MOE using its rule set in conjunction with Gasteiger [28] partial charges. Justification for the use of AMBER94 will be provided below; in brief, the choice of force field does not change the qualitative results. In order to maintain the integrity of the secondary structural elements of the polypeptide models in Fig. 1, the backbone atoms were held fixed in space; the possible effects of this constraint on CO_2 binding affinity will also be discussed below. Leucine was chosen as the side chain for the model polypeptides, as it can serve as an ideal control amino acid due to both its lack of functionality and its experimentally determined ability to be incorporated into either α-helices or β-sheets [25]. The side-chain residues were left unconstrained and minimized until the root-mean-square (RMS) gradient fell below 10^{-4} kcal mol^{-1} A^{-1}, first with a steepest descent algorithm until the RMS gradient was less than 1000 kcal mol^{-1} A^{-1}, then with a conjugate gradient algorithm until the RMS gradient was less than 100 kcal mol^{-1} A^{-1}, and finally with a truncated Newton algorithm. A single molecule of CO_2 was then randomly added to the system (early tests showed that the starting location of CO_2 did not affect the results) and the entire system was again reoptimized. Finally, to prevent diffusion of CO_2 away from the polypeptide, a restraining rectangular wall was centered on the center of mass of the polypeptide and extended by 6Å beyond the most distant polypeptide atoms in the x, y, and z directions, followed by a final energy minimization procedure (which did not change the structure). For simulations with water, the water soak algorithm in MOE was used to solvate a polypeptide centered in the rectangular wall just mentioned with TIP3P water, followed by another round of energy minimization as above. For simulations without water, the distance-dependent dielectric model in MOE was used for implicit solvation; although simple, this approach should provide reasonable results for the uncomplicated electrostatics of the systems considered herein. The MD simulations were run in the NVT ensemble using the Nosé–Poincaré–Anderson algorithm [29] to solve the equations of motion. The system was heated from 0 K to 500 K in 1 ps; early simulations at 300 K led to trajectories where CO_2 would proceed to the nearest electrostatic interaction site and stay there throughout the simulation, whereas the elevated temperature allowed the sampling of all areas of the polypeptide models. Simulations were run for 10 ns with a 0.5 fs timestep, with atomic positions collected every 0.5 ps. Bond lengths to hydrogen atoms were kept constant.

In order to quantitatively compare the numerous MD simulations generated throughout this work, distances between the carbon of CO_2 and the backbone polypeptide nitrogens were measured (at 0.5 ps intervals) for each trajectory (yielding a total of 20,000 distances per poly-peptide nitrogen per trajectory). These distances were then categorized according to residue type (i.e., middle, end, edge, and cap; see Fig. 1), and a minimum distance at each time step was separately determined for each of these residue types. Minimum distance histograms for each residue type were then generated for each simulation. For example, in Fig. 2, the minimum distance histograms for the interaction of CO_2 with the edge nitrogens of a β-sheet are given for MD simulations generated using (a) AMBER94—the force field used for the bulk of this work, (b) AMBER99 [30], (c) CHARMM27 [31–33], and (d) TraPPE [34]—a force field optimized to properly describe the properties of bulk CO_2. The significance of the peaks centered just above 4Å will be discussed in great detail later, but of relevance to the current discussion is the fact that these peaks, as well as the smaller peak between 5 and 6Å, are all the same shape, and thus yield the same qualitative picture; as AMBER94 yields a peak height intermediate between AMBER99 and CHARMM27, it is selected for use as the force field giving the most average or typical results. The plot in Fig. 2d is smaller because in TraPPE the charges on the C and O atoms in CO_2 are assigned values of $+0.70e$ and $-0.35e$, respectively, whereas the rule set implemented in MOE assigned Gasteiger charges of $+0.97e$ and $-0.49e$, respectively, for these values. Although the partial charges utilized in TraPPE lead to faithful reproductions of experimental liquid–vapor equilibrium curves, there are indications that larger partial charges may be more appropriate. For example, based on an MP2/6-31 G* quantum mechanical calculation, the partial charge on carbon in CO_2 is $+0.92e$ using the CHelpG [35] approach, $+0.86e$ using the ESP approach [36], and $+0.92e$ using the Mulliken population analysis scheme [37]. The Lennard–Jones parameters set by MOE for CO_2 are $\sigma=2.96$Å and $\varepsilon=0.21$ kcal mol^{-1} for O and $\sigma=3.40$Å and $\varepsilon=0.086$ kcal mol^{-1} for C, although previous MD simulations of CO_2–H_2O systems have indicated that results are generally insensitive to these values [38]. In any event, the fact that all peaks in Fig. 2 have the same shape indicate that the choice of force field may have a quantitative impact, but will not change the qualitative behavior described throughout this paper.

Another possible cause of error in these simulations is the fixing of the polypeptide backbone atoms, and a number of additional test MD simulations were performed to judge this effect. Figure E in the "Electronic supplementary material" (ESM) shows the minimum distance histograms for the doubly substituted Arg β-sheet model where one Leu residue was present between the two Arg residues (see Fig. 8); Fig. Ea illustrates CO_2 binding to the edge residues, and Fig. Eb shows binding to the nitrogen atoms of the Arg side chains. Figures Ec and d, likewise, show the metrics for a separate MD simulation where one of the Arg

Fig. 2 Histograms of minimum distance (in Å) from CO_2 to the edge nitrogen atoms of an all-Leu β-sheet using **a** AMBER94, **b** AMBER99, **c** CHARMM27, and **d** the TraPPE force field
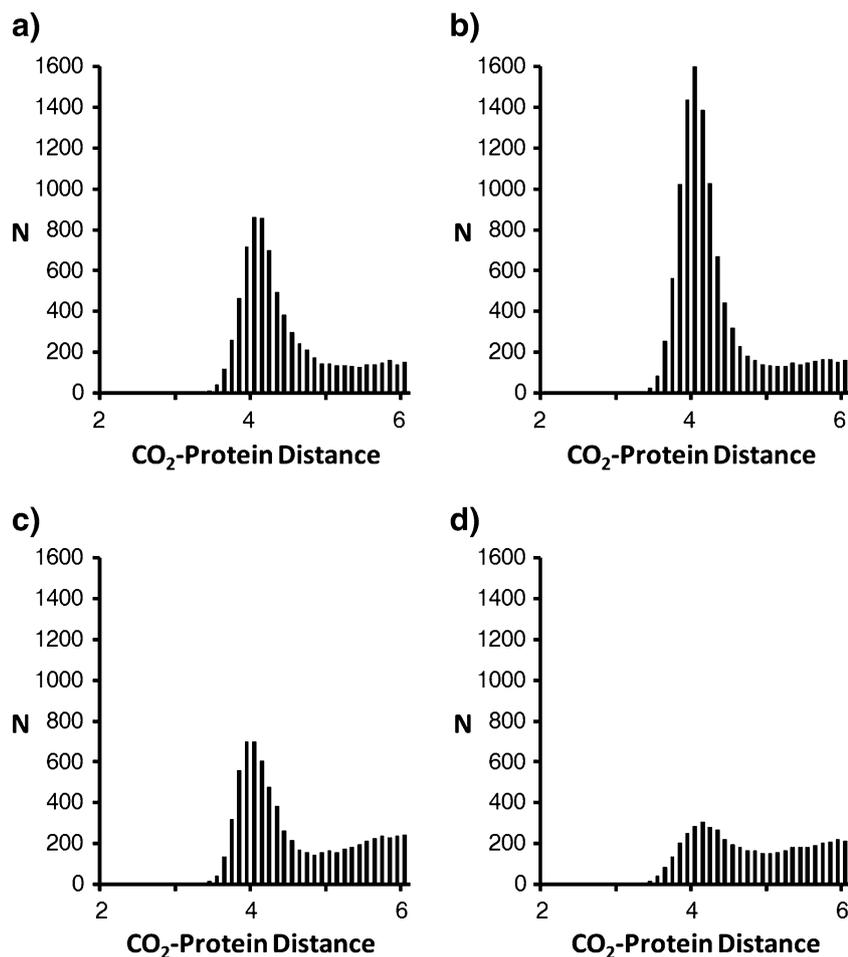
side chains was held fixed. It can be seen that, although Arg binding increases significantly due to fixing the position of one of the side chains (cf. Figs. Eb and d), binding to the edge of the polypeptide is not significantly affected (cf. Figs. Ea and c). Indeed, Fig. F in the ESM is parallel to Fig. 10, except that all Arg side chains have been fixed in space. As a result, binding of CO_2 to the edge residues increases, decreases, decreases, and increases for the situations where multiple Arg substitutions were made to the middle, cap, edge, and all residues, respectively. Finally, Fig. G in the ESM shows CO_2 binding to the edge residues along two separate all-Leu β-sheets: one (Fig. Ga) where the middle four residues of each β-strand in Fig. 1 were not fixed, and one (Fig. Gb) where the outer six residues of each β-strand in Fig. 1 were similarly left free to move. These two histograms can be compared to Fig. 3c, showing that relaxing the constraints on the backbone residues does diminish edge binding, but not tremendously so. These three figures in the ESM collectively indicate that, although fixing the backbone residues in space does enhance their CO_2 affinity somewhat relative to completely free residues, the difference is not large enough to change the conclusions of

this work, especially when it is considered that backbone atoms generally will indeed be held in place more rigidly in a full protein than will the more flexible side-chain residues.

Results and discussion

A comparison of CO_2 interactions in α-helices and β-sheets

Figure 3 shows a collection of minimum distance histograms comparing CO_2 affinity between an all-Leu α-helix and an analogous β-sheet. Additional trajectories were generated for the two secondary structures, differing only in the initial location of CO_2, but the histograms (not shown) are nearly identical to those in Fig. 3. The largest peak in Fig. 3 is in 3a, which corresponds to hydrogen bonding between CO_2 and the backbone amide groups of the edge residues of the β-sheet. This peak is not due to CO_2 binding to merely one edge amide group, but instead to all edge residues that have the –NH hydrogen bond donor oriented outward (and thus lacking a polypeptide partner acceptor) rather than inward towards the C=O hydrogen-bond

Fig. 3 Histograms of minimum distance (in Å) from CO_2 to **a** the edge, **c** the cap, and **d** the middle residues of an all-Leu β-sheet, and **b** the end and **e** the middle residues of an all-Leu α-helix. See Fig. 1 for an explanation of the residue types
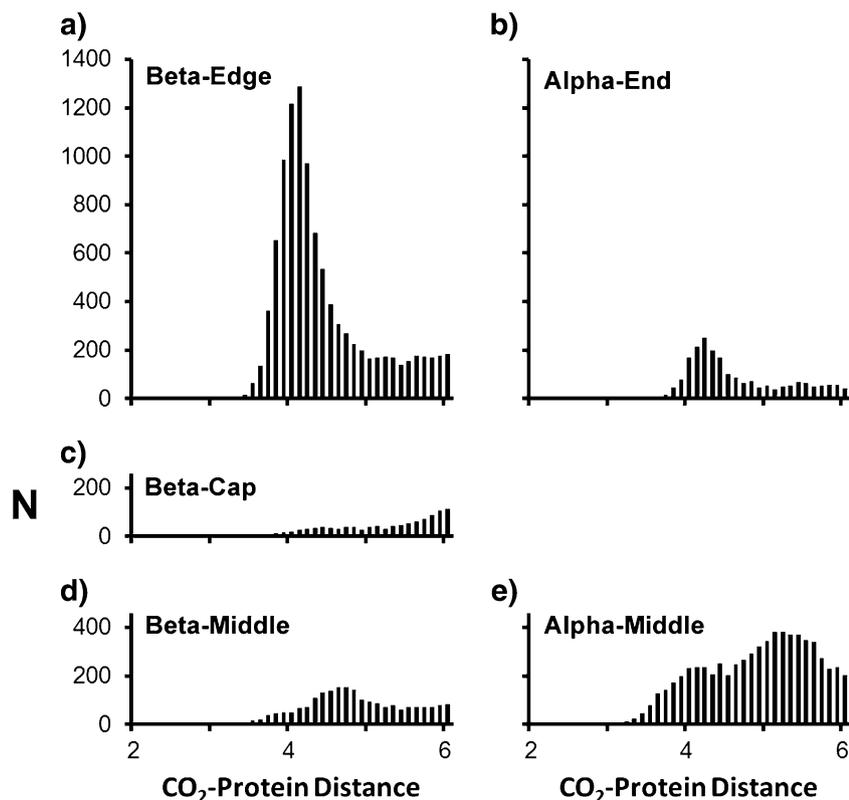
acceptor of a neighboring β-strand. This protein–NH⋯O=C=O hydrogen bond is representative of the most common hydrogen-bond type identified in a survey of various protein–ligand complexes [39].

This peak in Fig. 3a due to CO_2 binding to edge residues in the β-sheet is far larger than its analog in Fig. 3b due to CO_2 binding to the end residues in the α-helix. In a sense, this comparison seems uneven, given that there are only two end residues with unfulfilled –NH hydrogen-bond donors, compared to ten outwardly oriented, available –NH donors among the 20 edge residues. However, in another sense, this disparity is an important point: a β-sheet will always have amide hydrogen bond donors (and carbonyl acceptors) along its edge that can potentially bind a ligand, whereas an α-helix will have similarly exposed hydrogen-bond centers only at its ends (if at all). A parallel observation has been previously offered by Bartlett et al. [16], who attributed an "underrepresentation" of α-helices in catalytic sites to the fact that the hydrogen-bonding potential of the backbone atoms is completely satisfied through intrahelical hydrogen bonds, in direct contrast to the edges of β-sheets. Similarly, in a study of the location of buried water molecules in X-ray protein structures [40], hydrogen bonding was again invoked to explain the disparities in the abilities of the various secondary structure types to bind water. Additional evidence will be presented below to further explore the role of secondary structural elements, in particular the edges of β-sheets, in determining CO_2–protein binding patterns.

First, however, one final interesting difference in Fig. 3 can be seen between 3d and e, where the minimum distances between CO_2 and the middle residues of the β-sheet and the α-helix, respectively, are given. While neither possesses an especially large peak, there is evidence of greater CO_2 affinity for the middle residues of the α-helix than in the β-sheet. Just as in the discussion above, this difference is not merely a consequence of the greater number of residues in the former, but instead possesses a physical origin, as shown in Fig. 4. The electrostatic potential of the all-Leu α-helix and β-sheet are mapped onto the Connolly surface [41, 42] and depicted in Fig. 4b and d, while 4a and c show all-Gly analogs; the black dots correspond to the position of the carbon atom of CO_2 at each 0.5 ps timestep in the trajectory. For both secondary structural elements, the smallest amino acid, Gly, leads to considerable exposure of the electrostatic character of the backbone atoms, and thus CO_2 binds to the middle residues in both the α-helix of Fig. 4a and the β-sheet of Fig. 4c. With the larger Leu side chain, however, only the α-helix shows exposed backbone electrostatic character (as has been noted elsewhere [15]), which leads to localized CO_2 binding in the grooves of the helix (which is responsible for the peak in Fig. 3e), despite the absence of any available hydrogen-bonding centers in the middle of the α-helix.

Fig. 4 Electrostatic surfaces mapped onto the Connolly surface of **a** an all-Gly α-helix, **b** an all-Leu α-helix, **c** an all-Gly β-sheet, and **d** an all-Leu β sheet. *Red* corresponds to regions of partial negative charge, *blue* corresponds to regions of partial positive charge, and *green* corresponds to neutral regions. Each *black dot* corresponds to the location of the carbon atom of CO_2 (at 0.5 ps intervals)
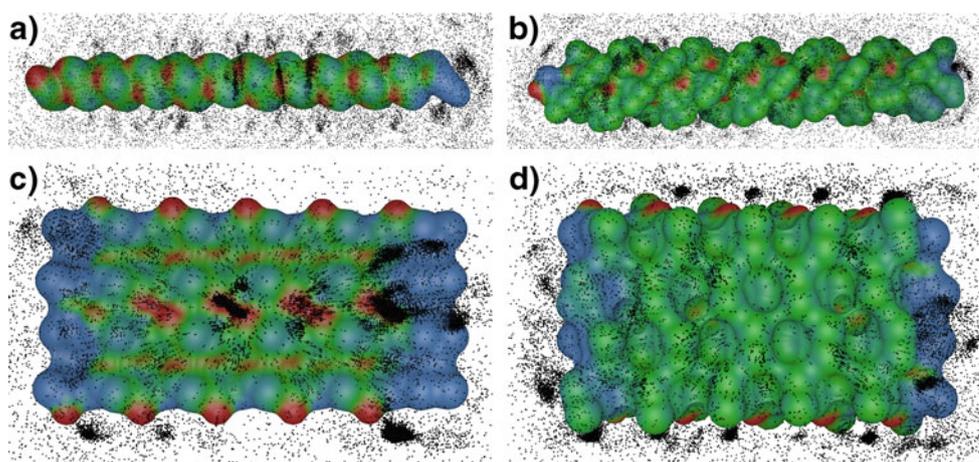
Thus, in addition to the $-NH \cdots O=C=O$ hydrogen bonding discussed above for CO_2 ligation by the edge residues, these polypeptides can also bind CO_2 electrostatically, although the interaction is weaker (cf. the greater peak in Fig. 3a vs. e).

The role of electrostatics in CO_2 binding to secondary structural elements

In order to more fully explore the balance between hydrogen bonding and a more generalized electrostatic interaction in determining protein–CO_2 affinity, the charges of the CO_2 molecule were manipulated in silico in two ways. First, all carbon dioxide partial charges were set to zero (see Fig. A, ESM). With only weak van der Waals attraction operative in this scenario, no direct binding of CO_2 is seen (indeed, MD simulations of protein–N_2 and protein–O_2 systems, not shown here, yielded results nearly identical to those in Fig. A). Secondly, and more interestingly, the signs of the partial charges on the atoms of CO_2 were reversed; that is, C was changed from its AMBER94 [27] charge of +0.974 to −0.974, and each O was similarly changed from −0.487 to +0.487. The minimum distance histograms for these simulations are given in Fig. 5.

Both end and edge residues show diminished CO_2 binding when the charges are reversed than with the usual charges (cf. Fig. 5a and b with Fig. 3a and b). This decrease can be attributed to an absence of backbone–CO_2 hydrogen bonding in the reversed charge situation: the end and edge backbone –NH amide groups can no longer donate a hydrogen bond to the oxygen of CO_2, as it has been assigned a δ^+ charge; moreover, the δ^- carbon lacks the requisite lone pair to act as a hydrogen-bond acceptor. Thus, the peaks seen in Fig. 5a and b can be attributed to electrostatic interactions. Visual inspection of these trajectories reveals that, rather than the end-on hydrogen-bonded situation exhibited by the trajectories of normally charged CO_2, the reversed charge CO_2 binds to the polypeptides in

a side-on fashion, where the partially negative backbone carbonyl is directed towards the partially positive oxygen of CO_2 and the partially positive amide hydrogen is directed towards the partially negative carbon.

That this side-on, electrostatic interaction is weaker than an end-on hydrogen bond is depicted in both the diminished size and increased width (reflecting increased flexibility) of the peaks in Fig. 5a and b as compared to their partners in Fig. 3. To estimate the strength of a typical CO_2–backbone hydrogen bond, the binding energy of CO_2 to N-methylacetamide (which has previously been suggested as an appropriate backbone model) [43] was calculated using the highly accurate correlation-consistent composite approach (ccCA) [23, 44, 45], which yields a CO_2–backbone hydrogen-bond strength of 1.5 kcal mol^{-1}, in line with the strength of an intrapeptide hydrogen bond (1–2 kcal/mol) [2]. The side-on, purely electrostatic interaction is weaker still, but it can nevertheless influence CO_2 affinity among the two secondary structure types. Indeed, Fig. 5e shows increased binding of CO_2 as compared to Fig. 3e, because the reversed charge CO_2 possesses two partially positive centers, and can thus straddle two partially negative backbone segments (see Fig. 4b).

Synthetic techniques to reduce CO_2 binding affinity along the edges of β-sheets

While the reversal of the partial charges on CO_2 proved useful in illuminating the roles of hydrogen-bonding and electrostatic interactions in the binding of CO_2 to polypeptides, it is unfortunately a solely in silico technique. However, Fig. 6 illustrates that two laboratory techniques which have proven useful in ameliorating the binding affinity of β-sheet edges—methylation [12] and esterification [11] of backbone amide groups—yield similar results. Replacing a backbone amide hydrogen with a methyl group destroys its hydrogen-bond donating capability, and thus

Fig. 5 Histograms of minimum distance (in Å) from reversed charge CO_2 (see text) to **a** the edge, **c** the cap, and **d** the middle residues of an all-Leu β-sheet and **b** the end and **e** the middle residues of an all-Leu α-helix. See Fig. 1 for an explanation of the residue types
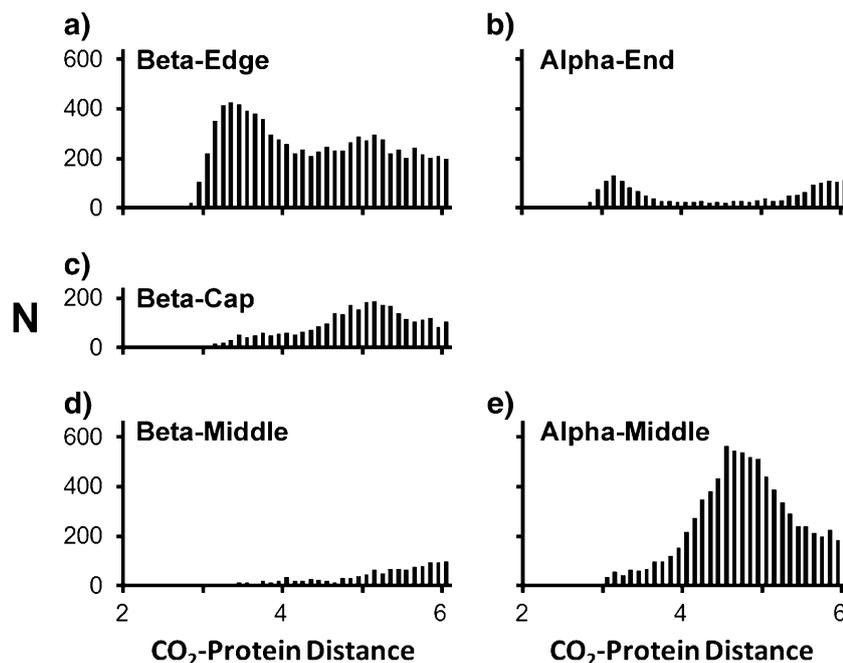
CO_2 binds to the edge of the β-sheet (Fig. 6a) in a side-on fashion, with the carbon of CO_2 bound by the backbone carbonyl groups.[1] There is also appreciable CO_2 binding by the cap residues (Fig. 6b), which possess a generally partial positive charge (Fig. 4) due to the capping methyl groups neighboring nitrogen and carbonyl moieties. While in a true protein the cap residues represent the point of transition from the β-sheet to the rest of the protein and thus would not be as sterically open to ligation as depicted in Fig. 4, the peak in Fig. 6b does illustrate once again that electrostatics can play a secondary role in CO_2 affinity, especially in the absence of available hydrogen-bonding centers.

Similar results are seen for esterification of the backbone amide groups: binding to the edge residues is diminished with respect to the unmodified β-sheet (Fig. 3a) but not removed entirely, as electrostatic attraction is still possible. Replacing only four of the 20 edge backbone amide groups—as was done experimentally [11]—yields distance metrics (not shown) that are intermediate between these two extremes. In addition, an esterified α-helix was also simulated (Fig. B of the ESM). As in the case of the β-sheet (Fig. 6c and d), this modification removes hydrogen bonding as a potential means to bind CO_2, and thus the end residues show very little CO_2 affinity. The

esterified α-helix still exhibits CO_2 binding through its middle residues, albeit at a slightly reduced level compared to that shown by the unmodified α-helix (Fig. 3e). This reduction can perhaps be attributed to increased electrostatic repulsion between the partially negative oxygen atoms of CO_2 and the more negative backbone (Fig. Bc of the ESM) in the esterified α-helix.

The effect on CO_2 affinity of adding a functionalized amino acid side chain

In addition to these two chemical techniques that are utilized to reduce β-sheet edge backbone interactions, an even more obvious effect on backbone interactions can be achieved by considering the incorporation of functionalized amino acid side chains rather than aliphatic Leu. In our previous bioinformatics-driven approach [23], we found that carbon dioxide is most commonly found in protein–CO_2 crystal structures near arginine. Indeed, the binding energy between the side chain of Arg and CO_2 was found, using ccCA, to be 6.1 kcal/mol, roughly four times stronger than the hydrogen bond (see above) between CO_2 and the backbone mimic N-methylacetamide.

Given this difference in hydrogen-bonding affinities, the presence of a single Arg residue (incorporated at the most central residue in both α and β polypeptide models) should have a noticeable effect on the polypeptide–CO_2 binding patterns, as a comparison of Fig. 7 with Fig. 3 reveals. The two end residue histograms (Figs. 3b and 7b) actually show increased affinity at the termini in the singly substituted Arg situation; in contrast, there are no significant differences

---

[1] The peak in Fig. 5a is actually shifted relative to the corresponding reversed charge peak in Fig. 4a, because the reversed charge carbon was partially negative, and thus attracted to the backbone amide group from which the distances are measured; the accurately charged carbon in Fig. 5a is partially positive and thus attracted by the backbone carbonyl group, which is one bond removed from the backbone amide group and thus more distant.

Fig. 6 Histograms of minimum distance (in Å) from CO_2 for **a** the edge and **b** the cap residues of a methylated β-sheet and **c** the edge and **d** the cap residues of an esterified β-sheet. Distances are measured from the backbone amide nitrogen in **a** and **b** and from the oxygen replacing the backbone amide nitrogen in **c** and **d**. See Fig. 1 for an explanation of the residue types
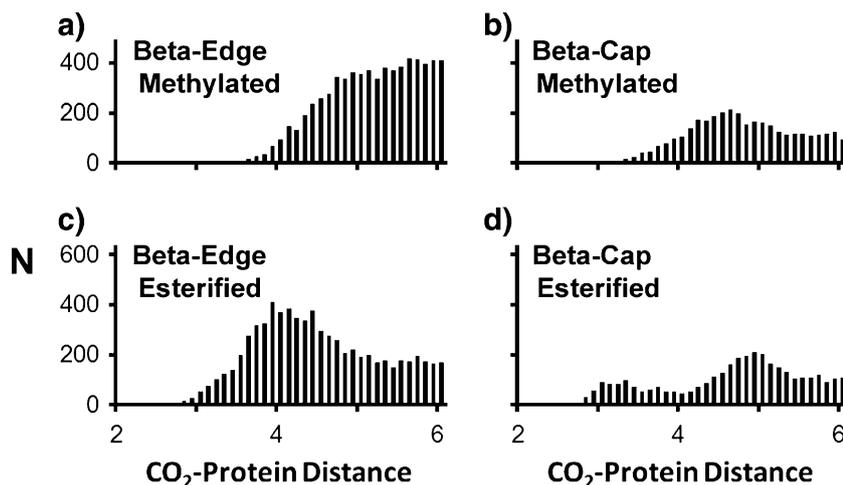
evident among the middle residues (Figs. 3e and 7f). Binding of CO_2 to the Arg side chain is somewhat greater in the α-helix than in the β-sheet (Fig. 7d and c, respectively); evidence will be presented below which indicates that this difference, albeit slight, is likely due to competition for CO_2 binding between the edge residues (Fig. 7a) and the Arg side chain in the case of the β-sheet. In any event, the edge binding seen in the singly substituted Arg β-sheet is decreased compared to the all-Leu simulation (Fig. 3a), but there is still significant edge binding. It is thus appropriate to see if there is further reduction of the edge–CO_2 interaction as additional Arg residues are incorporated into the primarily Leu polypeptide models.

The effect of adding two functionalized amino acid side chains on CO_2 affinity

Two Arg residues were systematically incorporated into the α-helix by first placing one in the middle of the helix and then adding a second Arg, first at a neighboring residue and continuing to increasingly distant residues, to a maximum Arg–Arg separation of 13 residues (which corresponds to, on average, >20Å of separation between these two side chains). For the β-sheet, the substitution pattern is less straightforward and is depicted in Fig. C (ESM). Substitutions were made to Leu residue pairs oriented in the same direction with respect to the plane of the β-sheet, and

Fig. 7 Histograms of minimum distance (in Å) for the singly substituted Arg polypeptide models: **a** the edge, **c** the Arg, and **e** the middle residues of a β-sheet, and **b** the end, **d** the Arg, and **f** the middle residues of an α-helix. Distances are measured from the amide nitrogen atoms to CO_2 for **a**, **b**, **e**, and **f**, and from the side-chain nitrogen atoms to CO_2 in **c** and **d**. See Fig. 1 for an explanation of the residue types
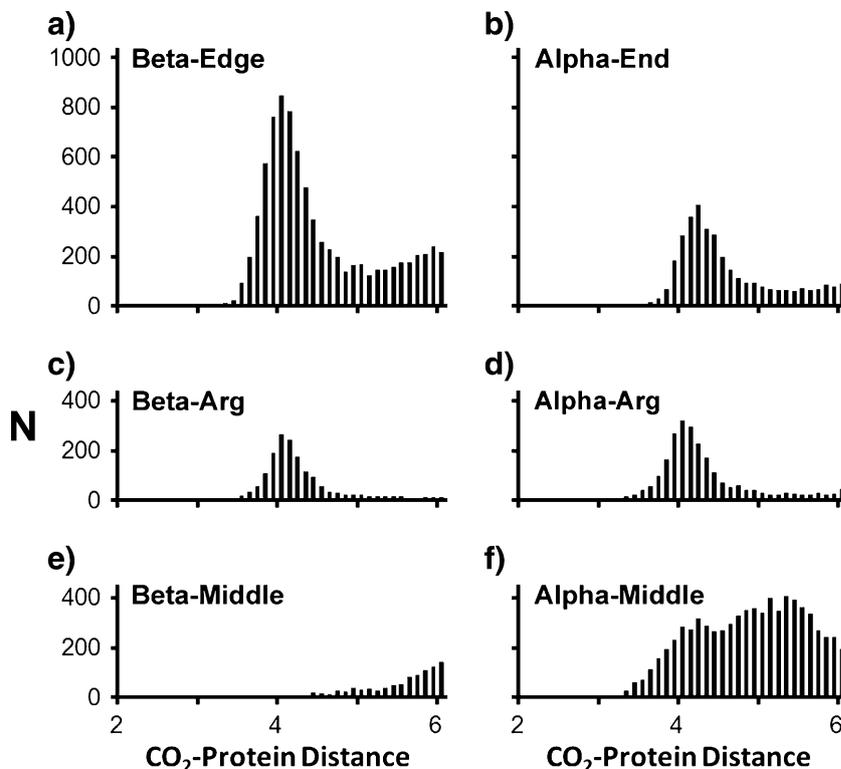
were only made to middle residues; the sole exception is in the case of the most distantly separated Arg–Arg simulation, which was performed on a polypeptide where one edge Leu above the plane and one edge Leu below the plane of the β-sheet were replaced by Arg. In addition, the entire suite of simulations was repeated with neutral His (with the hydrogen located on the δ-nitrogen) to examine the effect of inter-residue electrostatic repulsion between two protonated Arg side chains on CO_2 affinity.

For each doubly substituted Arg and doubly substituted His simulation performed, minimum distance histograms were created as before. All distances ≤5Å for each side chain–CO_2 histogram (as seen, for example, in Fig. 7c or d) were summed, indicating the ability of the two functionalized residues to directly bind CO_2 throughout a trajectory, and these are plotted as a function of residue separation in Fig. 8. At first glance, it seems that there may be evidence for anticooperativity between pairs of His residues in the β-sheet, which shows increasing CO_2 affinity as the two functionalized residues are increasingly separated. However, the other three sets of data in Fig. 8 do not support this trend; indeed, no trends are evident with respect to inter-residue distance whatsoever. It does appear, however, that Arg consistently shows a greater affinity for CO_2 than His does (as was revealed to be the case in protein–CO_2 X-ray crystal structures) [23]. Moreover, as in the singly substituted simulations (Fig. 7), Arg pairs located on the α-helix generally show a greater affinity for CO_2 than do the equivalent arginines on the β-sheet, likely due to the competition for CO_2 binding between the edge backbone hydrogen-bond donors and the Arg side chains.

The ≤5Å sums for the singly substituted Arg residues are 1499 for the β-sheet (Fig. 7c) and 2102 for the α-helix (Fig. 7d), respectively. In the case of the former secondary structure type, all doubly substituted Arg simulations lead to greater CO_2–Arg affinity than in the singly substituted Arg case, although the effect of the adding a second Arg was found to be minimal when the pairs were separated by one and four Leu residues. Indeed, if "cooperation" between two Arg residues with respect to binding CO_2 is defined as any CO_2 binding affinity above twice the contribution of a single Arg residue (1499), then only those β-sheet simulations where the Arg residues were separated by two and nine residues exhibit cooperation; in the latter case, there is clearly too much inter-residue separation for there to be direct Arg–Arg cooperation in binding CO_2. However, in the case of the α-helix simulations, evaluating the simulation where the pair of Arg residues were separated by two Leu residues—the only possible cooperative scenario in the β-sheet—instead gives evidence of *anti*-cooperation, as the two Arg residues show significantly reduced affinity for CO_2 as compared to the single Arg simulation depicted in Fig. 7d. Further complicating the analysis of Arg–Arg cooperation in the α-
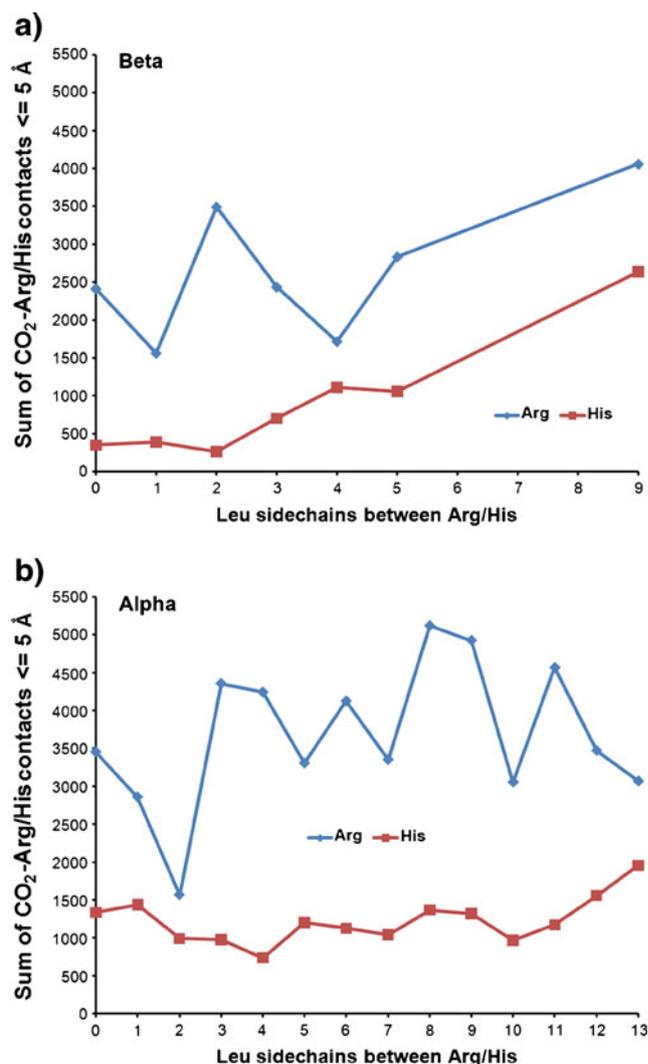
Fig. 8 The sum of all distances ≤5Å between CO_2 and the side-chain nitrogen atoms of Arg (*blue line*) or His (*red line*) for a given trajectory, plotted as a function of inter-residue separation, for **a** the β-sheet and **b** the α-helix. Refer to the text and ESM for details on the location of the functionalized residues

helix simulations, inter-residue separations of 3, 4, 8, 9, and 11 all exhibit cooperation as defined above, although only the first two simulations possess Arg residues that are close enough to cooperate and bind a molecule of CO_2 simultaneously. The conclusion, therefore, can only be that evidence for Arg–Arg cooperation (or anti-cooperation) in binding CO_2 is unclear, based on the results of these MD simulations.

The effect of the β-sheet edge residues on Arg–CO_2 binding, in contrast, is much clearer, as can be evaluated by considering Fig. 9, which is similar to Fig. 8a but instead considers CO_2–edge distances of ≤5Å rather than CO_2–Arg side-chain distances of ≤5Å. It was previously noted that Arg consistently exhibits greater CO_2 affinity than does His. However, as Fig. 9 reveals, this enhanced CO_2 binding exhibited by Arg relative to His does not necessarily come
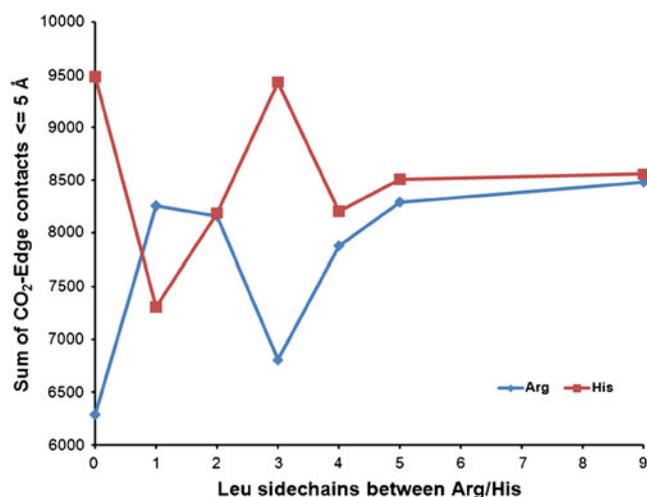
Fig. 9 The sum of all distances≤5Å between CO_2 and the β-sheet edge nitrogen atoms in the doubly substituted Arg (*blue line*) and His (*red line*) simulations, plotted as a function of inter-residue separation. Refer to the ESM for details on the location of the functionalized residues

at the expense of CO_2–edge residue binding. For example, in the simulations with one Leu residue between the pair of functionalized side chains, Arg binds CO_2 more than three times as often as does His (Fig. 8a). However, at this separation distance, more CO_2 is bound to the edge residues in the case of the doubly substituted Arg than in the doubly substituted His simulation. Similarly, although the two Arg residues bind CO_2 ten times as often as do a pair of His residues in the simulations where two Leu residues separate the functionalized side chains, the binding of CO_2 to the edge residues is essentially identical in the Arg and His simulations. If Arg were strictly competing with the edge residues for CO_2 binding, Fig. 9 would essentially be flipped with respect to the blue line in Fig. 8a; greater Arg affinity would, under this supposition, lead to diminished edge affinity, and vice versa. This is not so, however. Furthermore, in the singly substituted Arg simulation, there were 7229 occurrences of CO_2 within 5Å of the edge nitrogen atoms (Fig. 7a); only the doubly substituted simulations with zero and three Leu residues between Arg residues (Fig. 9) show decreased edge–CO_2 binding with two Arg than with one Arg. In other words, the addition of a second Arg residue, in general, seems to enhance the affinity of the edge residues for CO_2.

To summarize, it appears that the evidence for cooperation (or anti-cooperation) between two Arg residues in binding a molecule of CO_2 is mixed at best. In addition, Arg on an α-helix generally exhibits greater CO_2 binding than does Arg on a β-sheet, in both the singly and doubly substituted simulations, likely due to the absence of competing backbone hydrogen-bonding potential in the tightly coiled α-helix [16]. As expected [23], Arg consis-

tently shows greater affinity for CO_2 than does His. Perhaps most surprisingly, however, the presence of two Arg residues does not usually lead to decreased edge–CO_2 binding, but instead generally seems to lead to enhanced binding of CO_2 along the backbone of the edge residues. The limits of this last conclusion will be explored below through simulations of multiply substituted Arg β-sheets.

The effect on CO_2 affinity of incorporating multiple Arg residues into a β-sheet

Four MD trajectories were generated and analyzed to investigate the effect of multiple Arg residues on CO_2 binding patterns in β-sheets: three where every middle, cap, and edge Leu residue, respectively, were replaced by Arg, and a fourth where all Leu residues were replaced by Arg. All four simulations show increased Arg–CO_2 affinity with respect to both the singly (Fig. 7c) and doubly (Fig. 8) substituted Arg simulations; histograms illustrating Arg–CO_2 distances are not presented, but the total number of Arg–CO_2 distances of≤5Å for the four simulations are 17560, 7523, 15238, and 19143, respectively. Figure 10 presents the minimum distance histograms between CO_2 and the edge nitrogen atoms for the four simulations. While replacing the Leu side chains on the middle and cap residues with Arg residues reduces the amount of CO_2–edge binding (cf. Fig. 3a), the two simulations that possess Arg residues on the edges of the β-sheet (Fig. 10c and d[2]) do not show diminished CO_2 binding compared to the all-Leu situation. Indeed, the sum of CO_2–edge distances≤5Å in the multiply substituted edge simulation is 8660, greater than the total of 7229 observed in the singly substituted Arg simulation (Fig. 7a), and also greater than any of the totals seen in the doubly substituted Arg simulations (Fig. 9). Hence, the enhancement effect (see above) exerted by Arg residues on edge–CO_2 binding persists even as the number of Arg side chains is increased, despite the stronger interaction between CO_2 and Arg relative to that with the backbone (6.1 and 1.5 kcal/mol, respectively).

The obvious question, therefore, is why the Arg side chains do not always outcompete the backbone with respect to binding CO_2. One possible explanation is that the implicit solvation model (see "Computational methods") used in all of these simulations serves to reduce the disparity in binding energies, which were calculated with ccCA in a gas phase environment; this possibility will be addressed below in the context of considering explicit

---

[2] The size of the peak in Fig. 9d may be somewhat misleading, as approximately half of these distances correspond to CO_2 molecules bound by Arg residues above the plane of the β-sheet, rather than in an end-on fashion by backbone edge amide groups (see the ESM, Fig. D). Regardless, the general point about the enhancement effect of Arg on edge–CO_2 binding remains true, particularly as shown in Fig. 9c.

Fig. 10 Histograms of minimum distance (in Å) between the edge residue nitrogen atoms and CO_2 for the multiply substituted Arg β-sheet models, where the **a** middle, **b** cap, **c** edge, and **d** all Leu residues were replaced by Arg. See Fig. 1 for an explanation of the residue types
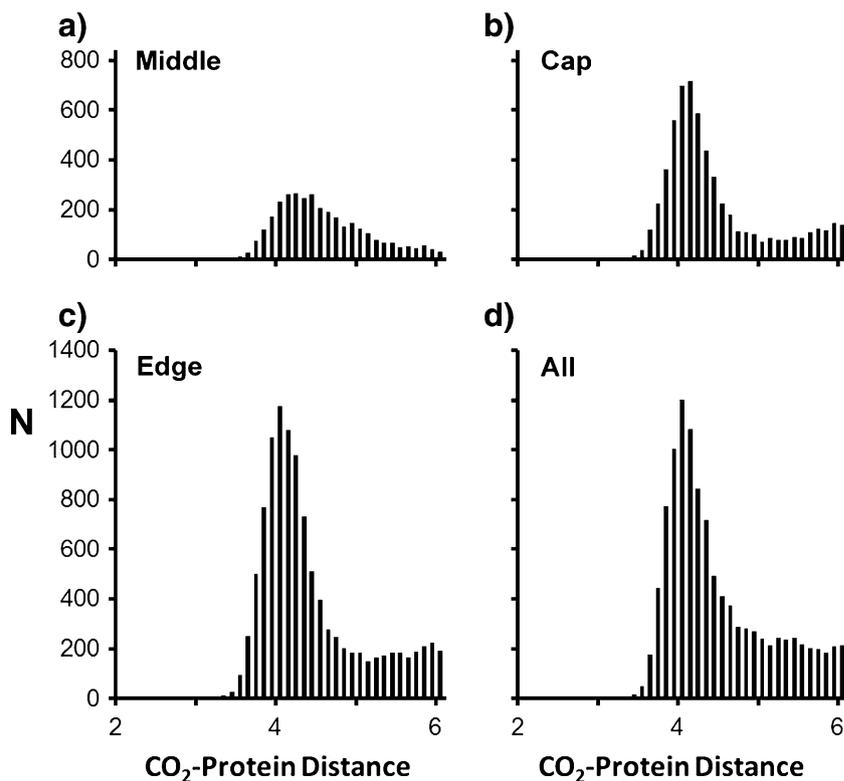
solvent MD simulations. However, regardless of whether an artifact of the computational procedure exists, we offer as conjecture the fact that the large number and relatively tight packing of the β-sheet edge interaction sites results in the ability of the backbone to successfully bind CO_2, even if multiple Arg residues are located nearby. Indeed, the situation depicted in Fig. 10 that most strongly reduced edge–CO_2 binding—Fig. 10a, where 16 Arg residues all neighbor one another in the middle of the β-sheet—is not likely to be found in any natural protein. In reality, the attractive influence exerted by Arg on CO_2 will be "watered

down" by steric interactions with nearby, nonfunctionalized side chains, such as Ala, Leu, Ile, Phe, etc. By contrast, the polypeptide model β-sheet backbone is quite similar to the backbone in a true protein, as proteins do indeed possess multiple amide groups, mostly uniformly oriented and capable of donating to a hydrogen bond. Thus, in effect, the concatenation of numerous, closely located (albeit weaker) backbone interaction sites is enough to effectively compete in binding CO_2 with the stronger but largely isolated Arg side-chain interaction sites. Hence, if one of the weaker backbone–CO_2 hydrogen bonds is broken due

Fig. 11 Histograms of minimum distance (in Å) for **a** the edge, **c** the cap, and **e** the middle residues of an all-Leu β-sheet in a box with 670 explicit water molecules, and **b** the edge, **d** the cap, and **f** the middle residues of a singly substituted Arg β-sheet in a box with 670 explicit water molecules. Distances are measured from the amide nitrogen atoms of the respective residue types (see Fig. 1) to CO_2
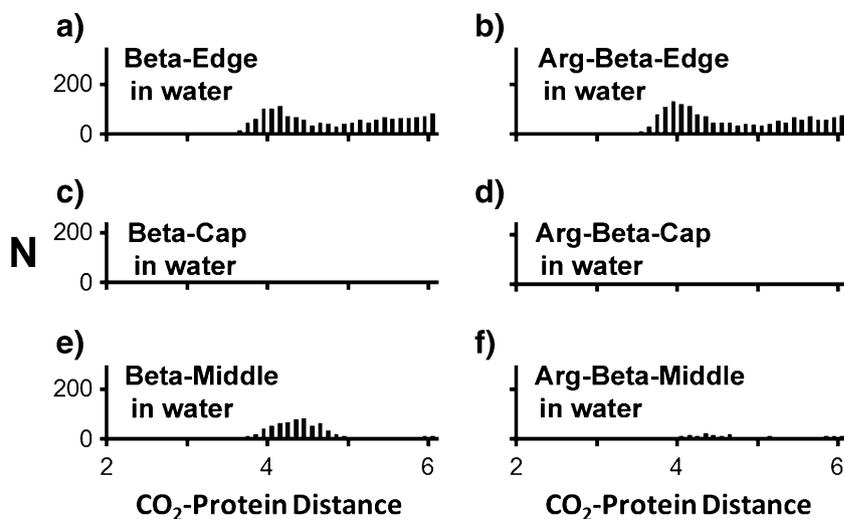
to vibrational motion, solvent interference, etc., other potential (if equally weak) hydrogen-bond donors are located nearby along the β-sheet backbone. The fact that numerous protein residues are appropriately "lined up" in ligand binding sites has previously been observed [39].

The effect of the water solvent on β-sheet affinity for CO_2

One final factor that could influence the ability of β-sheets to bind CO_2 is the solvent water molecules. While all results discussed up to this point have incorporated an implicit model to indirectly account for solvent effects, trajectories were also generated for both an all-Leu and a singly substituted Arg β-sheet with water molecules specifically included. Figure 11 presents the distance metrics for these two polypeptides in a box of 670 water molecules; the binding of carbon dioxide to Arg (i.e., a histogram analogous to Fig. 7c) is not plotted for the latter simulation, as there are only 42 observations of CO_2 within 5Å of the side chain of Arg (out of 20,000 total observations). This lack of CO_2–Arg affinity in the face of competition with H_2O–Arg binding is unsurprising, as the respective binding energies are 6.1 and 13.8 kcal/mol at the ccCA level. Immediately obvious in Fig. 11 is the fact that the explicit water molecules impede the binding of CO_2, irrespective of residue type (cf. Figs. 11a, c, and e with Fig. 3a, c, and d and Fig. 11b and f with Fig. 7a and e). Binding of CO_2 to the cap residues, rare to begin with, is completely eliminated, as the water molecules preferentially solvate this partially positive region (see Fig. 4d). CO_2 binding through the edge residues is also severely impacted, although the enhancement effect exerted by Arg on CO_2–edge binding (discussed in the previous two sections) is in evidence here as well: the sum of CO_2–edge interactions within 5Å is 881 for the all-Leu simulation (Fig. 11a), compared to 1032 for the singly substituted Arg simulation (Fig. 11b). While this difference is small, additional simulations performed with 220 water molecules (not shown) yield similar results: 2084 CO_2–edge contacts within 5Å for the all-Leu simulation, compared to 3239 for the singly substituted Arg simulation.

Conclusions

In contrast to the situation with larger ligands, such as ATP or a typical druglike molecule, the small size of CO_2 leads to the existence of many competing binding sites on a protein. In this study, a series of CO_2–polypeptide MD trajectories were generated to explore the nature of the possible interactions (i.e., hydrogen-bonding, electrostatic or van der Waals attraction), the relative strengths of these interactions with respect to polypeptide functionality (e.g., the binding of CO_2 to the edge residues of a β-sheet, or to

the middle residues of an α-helix, or to the Arg side chains, etc.), and the mutual influences exhibited by the various polypeptide functional groups on one another.

These MD simulations lead to the conclusion that β-sheets do indeed show greater affinity for CO_2 than α-helices, as we have previously observed [23], and that the cause of this disparity can be attributed to the presence of available amide hydrogen-bond donors along the edge of the β-sheet; the equivalent hydrogen-bonding centers are exclusively involved in intrapeptide hydrogen bonding in an α-helix. The strength of this CO_2–backbone hydrogen bond is comparable to the estimate [2] for a typical intrapeptide bond. Turning off the backbone hydrogen-bonding capability of the β-sheet, be it by computationally reversing the charges on C and O or by experimentally modifying the backbone amide groups through methylation or esterification, leads to a weaker electrostatic interaction between CO_2 and the β-sheet, again primarily along the edge. In most trajectories, introducing one or more functionalized amino acid side chains into the polypeptide, such as the guanidinium group of Arg, actually increases the occurrence of CO_2 binding along the β-sheet edge backbone, even in the presence of water.

The picture that emerges from these molecular dynamics simulations thus seems to be that, rather than strict competition between Arg– and edge–CO_2 binding, Arg can be responsible for drawing CO_2 into the vicinity of the hydrogen-bond donors along the β-sheet backbone. We are currently performing MD simulations on interactions between full proteins and CO_2 to explore this phenomenon on a larger scale and to expand the scope beyond Leu, Arg, and His to the entire suite of naturally occurring amino acids. At this point, however, it seems clear that the interplay between backbone and side-chain groups plays a critical role in determining binding patterns of CO_2, as seems to be the case in protein folding as well [3], and thus the results of this work will have important consequences in ongoing and future efforts to elucidate protein-based, environmentally friendly carbon capture systems.

References

1. Stickle DF, Presta LG, Dill KA, Rose GD (1992) Hydrogen bonding in globular proteins. J Mol Biol 226:1143–1159
2. Fleming PJ, Rose GD (2005) Do all backbone polar groups in proteins form hydrogen bonds? Protein Sci 14:1911–1917

3. Rose GD, Fleming PJ, Banavar JR, Maritan A (2006) A backbone-based theory of protein folding. Proc Natl Acad Sci USA 103:16623–16633

4. Lu W, Randal M, Kossiakoff A, Kent SBH (1999) Probing intermolecular backbone H-bonding in serine proteinase–protein inhibitor complexes. Chem Biol 6:419–427

5. Moreira IS, Fernandes PA, Ramos MJ (2007) Backbone importance for protein–protein binding. J Chem Theor Comput 3:885–893

6. Lazo ND, Maji SK, Fradinger EA, Bitan G, Teplow DB (2005) The amyloid beta-protein. In: Sipe JC (ed) Amyloid proteins—the beta sheet conformation and disease. Wiley-VCH, Weinheim, pp 385–492

7. Teplow DB, Lazo ND, Bitan G, Bernstein S, Wyttenbach T, Bowers MT, Baumketner A, Shea J, Urbanc B, Cruz L, Borreguero J, Stanley HE (2006) Elucidating amyloid β-protein folding and assembly: a multidisciplinary approach. Acc Chem Res 39:635–645

8. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366

9. Laidman J, Forse GJ, Yeates TO (2006) Conformational change and assembly through edge β-strands in transthyretin and other amyloid proteins. Acc Chem Res 39:576–583

10. Richardson JS, Richardson DC (2002) Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci USA 99:2754–2759

11. Gleitsman KR, Lester HA, Dougherty DA (2009) Probing the role of backbone hydrogen bonding in a critical β-sheet of the extracellular domain of a Cys-loop receptor. Chem Bio Chem 10:1385–1391

12. Doig AJ (1997) A three stranded β-sheet peptide in aqueous solution containing N-methyl amino acids to prevent aggregation. Chem Commun 2153–2154

13. Das A, Mukhopadhyay C (2009) Urea-mediated protein denaturation: a consensus view. J Phys Chem B 113:12816–12824

14. Hua L, Zhou R, Thirumalai D, Berne BJ (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. Proc Natl Acad Sci USA 105:16928–16933

15. Rossky PJ (2008) Protein denaturation by urea: slash and bond. Proc Natl Acad Sci USA 105:16825–16826

16. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. J Mol Biol 324:105–121

17. Martin AC, Orengo CA, Hutchinson EG, Jones S, Karmirantzou M, Laskowski RA, Mitchell JB, Taroni C, Thornton JM (1998) Protein folds and functions. Structure 6:875–884

18. Schneider G, Lindqvist Y, Brändèn C-I (1992) RUBISCO: structure and mechanism. Annu Rev Biophys Biomol Struct 21:119–143

19. Tcherkez GGB, Farquhar GD, Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc Natl Acad Sci USA 103:7246–7251

20. Bond GM, Stringer J, Brandvold DK, Simsek FA, Medina M-G, Egeland G (2001) Development of integrated system for biomimetic CO_2 sequestration using the enzyme carbonic anhydrase. Energy Fuel 15:309–316

21. Ozdemir E (2009) Biomimetic CO_2 sequestration. 1. Immobilization of carbonic anhydrase within polyurethane foam. Energy Fuel 23:5725–5730

22. Nelson DL, Cox MM (2005) Lehninger: principles of biochemistry. WH Freeman, New York

23. Cundari TR, Wilson AK, Drummond ML, Gonzalez HE, Jorgensen KR, Payne S, Braunfeld J, Jesus MD, Johnson VM (2009) CO_2-formatics: how do proteins bind carbon dioxide? J Chem Inf Model 49:2111–2115

24. Drummond ML, Wilson AK, Cundari TR (2010) Toward greener carbon capture technologies: a pharmacophore-based approach to predict CO_2 binding sites in proteins. Energy Fuel 24:1464–1470

25. Costantini S, Colonna G, Facchiano AM (2006) Amino acid propensities for secondary structures are influenced by the protein structural class. Biochem Biophys Res Comm 342:441–451

26. Chemical Computing Group (2008) Molecular Operating Environment (MOE). Chemical Computing Group, Montreal

27. Cornell WD, Cieplak P, Bayly CI, Gould IR, K. M. Merz J, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197

28. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 26:3219–3228

29. Bond SD, Benedict JL, Laird BB (1999) The Nosé–Poincaré method for constant temperature molecular dynamics. J Comput Phys 151:114–134

30. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21:1049–1074

31. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616

32. MacKerell AD Jr, Banavali N (2000) All-atom empirical force field for nucleic acids. 2. Application to molecular dynamics simulations of DNA and RNA in solution. J Comput Chem 21:105–120

33. Foloppe N, MacKerell AD Jr (2000) All-atom empirical force field for nucleic acids. 1. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 21:86–104

34. Potoff JJ, Siepmann JI (2001) Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. AICHE J 47:1676–1682

35. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials—the need for high sampling density in formamide conformational analysis. J Comput Chem 11:361–373

36. Besler BH, Merz KM Jr, Kollman PA (1990) Atomic charges derived from semiempirical methods. J Comput Chem 11:431–439

37. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GA (2000) A complete basis set model chemistry. VII. Use of the minimum population localization method. J Chem Phys 112:6532–6542

38. In Het Panhuis M, Patterson CH, Lynden-Bell RM (1998) A molecular dynamics study of carbon dioxide in water: diffusion, structure and thermodynamics. Mol Phys 94:963–972

39. Sarkhel S, Desiraju GR (2004) N–H…O, O–H…O, and C–H…O hydrogen bonds in protein–ligand complexes: strong and weak interactions in molecular recognition. Proteins Struct Func Bioinf 54:247–259

40. Park S, Saven JG (2005) Statistical and molecular dynamics studies of buried waters in globular proteins. Proteins Struct Func Bioinf 60:450–463

41. Connolly ML (1983) Solvent-accessible surfaces of proteins and nucleic acids. Science 221:709–713
42. Liang J, Edelsbrunner H, Fu P, Sudhakar P, Subramaniam S (1998) Analytical shape computation of macromolecules. I. Molecular area and volume through alpha shape. Proteins Struct Func Bioinf 33:1–17
43. Morozov AV, Kortemme T, Tsemekhman K, Baker D (2004) Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. Proc Natl Acad Sci USA 101:6946–6951
44. DeYonker NJ, Cundari TR, Wilson AK (2006) The correlation consistent composite approach (ccCA): an alternative to the Gaussian-n methods. J Chem Phys 124:114104
45. DeYonker NJ, Wilson BR, Pierpont AW, Cundari TR, Wilson AK (2009) Toward the intrinsic error of the correlation consistent composite approach (ccCA). Mol Phys 107:1107–1121