Published in final edited form as:

ACS Synth Biol. 2018 December 21; 7(12): 2908–2917. doi:10.1021/acssynbio.8b00455.

Phosphate Lock Residues of *Acidothermus cellulolyticus* Cas9 Are Critical to Its Substrate Specificity

Travis H. Hand¹, Anuska Das¹, Mitchell O. Roth¹, Chardasia L. Smith², Uriel L. Jean-Baptiste², and Hong Li^{1,2,*}

¹Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

²Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA

Abstract

Despite being utilized widely in genome sciences, CRISPR-Cas9 remains limited in achieving high fidelity in cleaving DNA. A better understanding of the molecular basis of Cas9 holds the key to improve Cas9-based tools. We employed direct evolution and in vitro characterizations to explore structural parameters that impact the specificity of the thermophilic Cas9 from *Acidothermus cellulolyticus* (AceCas9). By identifying variants that are able to cleave mismatched protospacers within the seed region, we found a critical role of the phosphate lock residues in substrate specificity in a manner that depends on their sizes and charges. Removal of the negative charge from the phosphate lock residues significantly decreases sensitivity to the guide-DNA mismatches. An increase in size of the substituted residues further reduces the sensitivity to mismatches at the first position of the protospacer. Our findings identify the phosphate lock residues as an important site for tuning the specificity and catalytic efficiency of Cas9.

Keywords

CRISPR-Cas9; Thermophilic Cas9; Cas9 off-target; Gene editing; Direct evolution

The clustered, regularly interspaced, short palindromic repeats (CRISPR) and the CRISPR-associated proteins (Cas) systems originate from bacteria and archaea, where they provide adaptive immunity against invading nucleic acids. The CRISPR-Cas immunity allows small pieces of foreign DNA from previous invaders to be incorporated into the CRISPR array as

*Correspondence to: hong.li@fsu.edu.

Author Contributions

T.H. and H.L. designed all experiments. T.H. designed primers for *ccdB* plasmid variants and AceCas9 plasmid libraries, created AceCas9 plasmid libraries, performed in vivo survival assays with AceCas9 libraries and mutants obtained from libraries, and performed in vitro DNA cleavage assays. A.D. purified AceCas9 and variants obtained from library screens. M.R., C.S., U.J.B. performed in vivo survival assays with wild-type AceCas9 on *ccdB* variants. T.H. and H.L. analyzed data, wrote and edited manuscript and figures.

Supporting Information

- Sequencing results of isolated constructs.
- Sequence alignment of published Cas9 structures phosphate loops with images.
- Secondary structure comparison of SauCas9 and CjeCas9 to the AceCas9 model.

Conflict of interest

The authors declare that they have no conflict of interest.

spacers which, when transcribed and processed, direct Cas proteins to destroy new intruding DNA or RNA^{1–31–31–31–3}. Cas9 is a Cas protein belonging to the Type II CRISPR-Cas system^{3–4} that functions with a CRISPR RNA (crRNA) and a trans-activating RNA (tracrRNA) in targeting and cleaving nucleic acid^{1,5–10}. The crRNA for Cas9 typically contains a guide segment, usually 20–24 nucleotides (nts) in length, at the 5' end that is complementary to target DNA (protospacer), as well as a repeat region at the 3' end that can pair with the antirepeat region of the tracrRNA (Figure 1). The repeat:anti-repeat stem, and the two to three stem loops (3' handle) within the tracrRNA are recognized by Cas9. The crRNA and tracrRNA can be covalently linked into a single-guide RNA (sgRNA) with little to no loss of function^{11–14} (Figure 1). The guide of the crRNA is “programmable”, thereby allowing Cas9 to target virtually any DNA sequences. In addition to the protospacer, the target DNA must also contain a Cas9 specific sequence called the protospacer adjacent motif (PAM) immediately downstream of the protospacer sequence in order for it to be cleaved by Cas9^{4,6–9} (Figure 1). The programmable cleavage of dsDNA both in vitro and in vivo by Cas9 has enabled the widespread development of Cas9-based genome biotechnology^{15–27}.

One limitation to a broad application of Cas9 is the off-target events. The best-studied Cas9 from *Streptococcus pyogenes* (SpyCas9) is known to bind and cleave at similar but undesired genomic locations at rates comparable to or higher than that of the perfectly matched site²⁸. Initial studies showed that these off-target events are particularly common when mismatches between the sgRNA and target DNA occur far from the PAM (PAM distal)²⁸. In cultured human cells, systematically altered sgRNAs with mismatches at each nucleotide in the target sequences showed that single mismatches, especially the ones near the PAM (PAM-proximal), reduced cleavage activity, but the magnitude of the effect was highly variable among the targets²⁸. Double mismatches were even less effectively cleaved, but the effect depends on the location as well as separation of the mismatches, and some sgRNAs carrying two mismatches maintained cleavage activity similar to that of perfectly matched targets^{28–34}. As many as five mismatches allowed cleavage by SpyCas9 in U2OS-EGFP cells and again, the location as well as separation of these mismatches determined the impact on cleavage activity²⁸. Based on the off-target studies both in vitro and in vivo, a number of useful empirical rules have been derived in an effort to aid the design of guides with minimal off-targets^{35–37}. However, the sequence-dependency of the off-target sensitivity remains difficult to explain and the specificity data derived from human cells should be interpreted with caution, as other factors such as cell types have been shown to impact detected off-targets²⁸.

Efforts to understand the biophysical basis for the off-targets have produced results in general agreement with those found in cells. Cas9 is essentially a single-turnover enzyme with DNA-binding/unwinding (R-loop formation) being the rate-limiting step^{32, 38–39}. As such, the energetic barrier for off-targets to dissociate from the Cas9-sgRNA complex may be higher than that of the chemical step, resulting in the more likely event of their cleavage⁴⁰. The increase in specificity of some Cas9 variants with weaker guide-target or Cas9-target interactions may be explained by their lowered energetic barriers for dissociation⁴⁰. Kinetic studies of a catalytically inactive Cas9 (dCas9) binding to a target containing systematically constructed single and double mismatches found that mismatches within the PAM adjacent, or the “seed”, region reduced the on-rate but, interestingly, double mismatches outside the

seed region could increase the off-rate (termed reversibility-determining region, or RDR), both of which would weaken dCas9 binding to the target. The kinetic effects of off-targets have been modeled by using probabilistic considerations that are in good agreement with experimental measurements on SpyCas9⁴¹. Though the kinetic parameters have provided an excellent framework for understanding the biophysics of off-targets, cellular environments can modify the relative energetic profile and thus perturb the kinetic model obtained in vitro. The ability to predict and ultimately eliminate off-targets of a Cas9 holds the key to more broad applications of Cas9 in genome editing.

Recent efforts to eliminate the undesired cleavage events of SpyCas9 have found increasing success. These include the use of truncated sgRNAs²⁸, paired nickases^{42–43}, chimeric proteins with inactive Cas9 fused to a non-specific FokI nuclease^{35,44}, and structure guided mutagenesis^{36–37,45–46}, among which, the structure guided mutagenesis seems to have lent the most improved specificity. For instance, enhanced specificity SpyCas9, eSpyCas9, improves specificity by 10-fold, as measured by targeted deep sequencing and unbiased whole-genome off-target analysis, over the wild-type and contains mutations, K848A, K1003A, R1060A, within the nucleic acid recognition (REC) domain that are believed to attenuate DNA-unwinding activity³⁷. High fidelity SpyCas9, SpyCas9-HF1, that rendered nearly all off-targets undetectable via genome wide-break capture and targeted sequencing methods, contains mutations N497A, R661A, Q695A, and Q926A also within the REC close to the bound DNA at the PAM distal end³⁶. Finally, hyper-accurate SpyCas9, HypaCas9, containing N692A, M694A, Q695A, H698A mutations improves specificity even more than the eSpyCas9 and SpyCas9-HF1 while retaining on target efficiency⁴⁷. These results suggest that Cas9 itself has a strong ability to influence the specificity and that the best prospect for improving the specificity lies within the interaction between Cas9 protein and the target DNA. In each of these successful examples, the mutated residues are largely mapped to sites that contact the target DNA, and were also altered to alanine, raising a possible correlation between the degree of SpyCas9 contacts with the target and the specificity.

We are interested in probing the rules governing specificity of the previously characterized thermophile Cas9 from *Acidothermus cellulolyticus* (AceCas9)⁴⁸, a Type II-C Cas9⁴⁹, by evolving its variants that recognize the otherwise deleterious protospacer mismatches in bacterial cells. We employed a directed evolution method based on co-expressing the toxic *ccdB* protein and an AceCas9 library similar to that used in DNA homing endonuclease evolution⁵⁰. In this method, survival of the bacterial cells transformed with both the *ccdB*- and AceCas9-encoding plasmids is linked to the activity of AceCas9. We found that tolerance to mismatches within the seed region of the protospacer by AceCas9 can be achieved by substituting the phosphate lock residues with bulky and non-negatively charged residues. The selected amino acids at this position against a mismatched protospacer provide an explanation for the critical role of the phosphate lock residues in maintaining high fidelity within the seed region.

Results

PAM-proximal mismatches are deleterious to AceCas9 activity in bacterial cells

Our previous in vitro DNA cleavage assays show that AceCas9 is sensitive to several protospacer mutations. Using a molar excessive of enzyme, plasmid DNAs bearing protospacer mismatches to the guide RNA at positions -1, -4, -8 (+1, +4, +8 of the target strand) were defective in cleavage, whereas those at positions -19, -20, and -24 (+19, +20, and +24 of the target strand) tolerated cleavage by AceCas9⁴⁸.

To test the protospacer sensitivity of AceCas9 in bacterial cells, we inserted four mutations at positions -1, -4, -20 and -24 (nontarget strand), respectively, to a protospacer on an arabinose (Ara)-inducible *ccdB*-encoding plasmid. Electrocompetent *Escherichia coli* (*E. coli*) cells harboring the wild-type and the mutant target plasmids were prepared before being transformed with a dual-expression vector encoding AceCas9 and its sgRNA. Successful cleavage of the *ccdB* plasmid by AceCas9 would yield cell growth in the presence of arabinose (Ara+), whereas inability for AceCas9 to cleave the target plasmid would result in cell death (Figure 2A). For each of the wild-type and mutant transformations, we computed a survival rate by dividing the number of colonies on the Ara+ plate by those on the Ara-plate. We found that the survival rates for mismatched protospacers at positions -1 and -4 were 0.18% and 0.20%, respectively, whereas those for the wild-type and mismatches at positions -20 and -24 were about 95% and 76%, respectively (Figure 2B). These results are consistent with our previous in vitro cleavage results and show that, like other Cas9s, AceCas9 has a strong dependence on the Watson-Crick base pairing between the target strand and sgRNA at PAM proximal positions, but not PAM-distal positions⁴⁵.

Phosphate lock variants tolerate -1 mismatch

The near complete death on Ara+ plates of the cells bearing the protospacer plasmid containing a mismatch at -1 nontarget position, T(-1)G, transformed with the wild-type AceCas9-encoding plasmid created an opportunity to search for AceCas9 variants that rescue this phenotype, thereby identifying AceCas9 residues responsible for sensitivity to the protospacer mismatch at this position. To investigate this possibility, we created several AceCas9 libraries containing amino acid variations in the PAM-Interacting domain (PID), the linker connecting RuvC to PID (RP-linker), and the Recognition (REC) domain (Figure 3A). We reasoned that these regions may have an impact on recognition of the base pair at +1 target position based on a constructed structural model of AceCas9 using I-TASSER⁵¹⁻⁵³ (Figure 1, 3B). Initial screens with either the REC or the PID libraries did not yield convincing rescue mutants for the T(-1)G mismatch. However, the focused RP-linker library containing randomized residues at positions 839-840, in contrast to the wild-type AceCas9, generated significant survival following the first round of transformation. Survivors were pooled and their plasmid DNA was extracted for subsequent rounds of survival and Next Generation Sequencing (NGS). The NGS reads from the extracted DNA from three rounds of survival as well as the input library were analyzed for the abundance of each of the 20 amino acids at positions 839 and 840 (Figure 3). In addition, we selected nine colonies from the final surviving pool for individual DNA extraction and sequencing. The plots of amino acid abundance at positions 839 and 840 in the final survival pool showed an enrichment of

arginine (R) at position 839 and tyrosine (Y) or leucine (L) at position 840 (Figure 3B). Consistently, each of the nine individually analyzed survival mutants contained an overrepresentation of RY (6 out of 9) with two RL and one RH at these two positions (Figure 3D, Figure S1), suggesting that mutation of the wild-type glutamate at position 839 (Glu839) to arginine is a key determinant for the survival of the mismatched plasmid at +1 target position. Additional mutation of the wild-type glutamate at position 840 (Glu840) to tyrosine further assists this role.

To confirm the observed rescue activities of the selected variants, plasmids encoding AceCas9 variants bearing arginine at position 839 and tyrosine, leucine, or histidine at position 840 (RY, RL, and RH) were generated and transformed into the cells harboring the T(−1)G plasmid. Consistently, these variants showed a significant increase in survival rate (39.4%, 16.4%, 7.3% for RY, RL and RH, respectively) compared to that of the wild-type AceCas9 (0.18%) (Figure 4A).

To further support the *in vivo* activity of RY, RL and RH, we carried out the *in vitro* DNA cleavage activity assay with both the wild-type and the T(−1)G plasmid substrates by the AceCas9 mutants. The mutant AceCas9s were purified similarly as the wild-type AceCas9 and combined with a transcribed sgRNA to form RNPs that were incubated with the two substrates at 50 °C for 10 minutes⁴⁸. Invariably, whereas the wild-type AceCas9 resulted in little cleavage, all three mutants cleaved the T(−1)G substrate completely under the same reaction condition. No difference in activity was observed on the perfectly matched substrate (Figure 4B).

In the AceCas9 structural model, residues 839–840 are mapped to near protospacer +1 target position and correspond to the previously identified phosphate lock residues in SpyCas9⁴⁵. The phosphate lock stabilizes the sharp turn (S-turn) formed by the phosphate backbone of the target strand. Glu840 would be directly under the +1 target nucleotide whereas Glu839 would point to the opposite direction into the minor groove of the base pairs 1–3 (Figure 3C). Similar to other Cas9s, the position corresponding to Glu840 is always occupied by a negatively charged residue (Figure S2). Substitutions of Glu839 with arginine and Glu840 with tyrosine/leucine/histidine in the survival AceCas9 mutants remove the negative charge at position 840 while potentially increasing the buried surface area of around the +1 nucleotide.

The sizes of phosphate lock residues impact sensitivity to mismatches

The fact that the RY/RL/RH AceCas9 mutants, but not the wild-type, cleaved the T(−1)G protospacer raises the possibility that the size of the phosphate lock residues influences AceCas9 sensitivity to the mismatch at the first position. We thus created an AceCas9 variant that contained a smaller residue at the same position, E840A (RA), and subjected it to the *in vitro* DNA cleavage and cell survival assays. The RA variant was able to survive the T(−1)G cell survival assay, although with a lower survival rate than that for RY and similar to those observed for RL and RH (Figure 4A). Furthermore, *in vitro* DNA cleavage assays showed a minimal cleavage of the T(−1)G plasmid but a strong cleavage of the wild-type plasmid (Figure 4B), supporting that a reduction in size at position 840 increases sensitivity of the enzyme to the mismatch at position +1.

Phosphate lock residues impact sensitivity to mismatches at other positions

We wondered if the phosphate lock had an impact on the fidelity of base pairing at other protospacer positions. We transformed the RY, RA, and the wild-type AceCas9 plasmids into the cells harboring the G(−4)A (nontarget), the Fk-AA, and the Fk-AAGC substrate plasmids (see notations in Figure 1A). G(−4)A contains a mismatch in the seed region whereas Fk-AA and Fk-AAGC contain mismatches in the most PAM-distal region (Figure 1A). As shown in Figure 5A, both RY and RA exhibited much greater survival rates than the wild-type in the G(−4)A cells, with little to no differences observed in survival of the Fk-AA and the Fk-AAGC cells. In vitro DNA cleavage activity on the three DNA substrates by the purified RY, RA, and the wildtype showed the same trend (Figure 5B). These results suggest that phosphate lock mutants that accommodate mismatches at the first position also tolerate mismatches at the fourth position but not those distal to the PAM. The fact that neither the wildtype nor any of the AceCas9 mutants rescued the Fk-AAGC cells indicate that the 20-nt complementarity between the target strand and the sgRNA is not sufficient for AceCas9 to cleave this substrate.

Buried surface area of the target is correlated to off-target sensitivity

The dependence of the T(−1)G sensitivity on phosphate lock residue sizes prompted us to examine the relationship between the target buried surface area and specificity more systematically. We computed the nucleotide-specific buried surface areas of the target strand by Cas9 for the four Cas9 whose crystal structures are available^{45, 54–57}. Figure 6 shows the plots of the buried surface areas as a function of nucleotide positions for each of the four structures.

Positions +2 to +4 of the targets in all four Cas9 have high solvent accessibility or low buried surface areas (35–45 Å²) (Figure 6). Consistently, this region is very sensitive to mismatches as shown for SpyCas9^{28, 58–59} and SauCas9¹⁴ and other Cas9^{57, 60} and falls within the empirically defined “seed” region. In contrast, position +1, which also belongs to the seed region, is significantly buried primarily through its sugar-phosphate backbone atoms by the phosphate lock residues whereas the reported sensitivity to mismatch is also high in most cases⁴⁵, suggesting that buried solvent area alone is not a good predictor of sensitivity to mismatches at this position. Interestingly, a second region, +13 –+15, outside the “seed” region, also exhibits relatively high solvent accessibility that indicates possible sensitivity to mismatches. Indeed, this region (+8 –+17) is consistent with the previously identified “reversibility-determining region (RDR)”³² that was observed to exclusively modulate the off-rate of DNA binding. Single mismatches within RDR, when combined with mismatches at the most PAM-distal end (+19 or +20) where single mismatches produced nearly no change, were observed to noticeably reduce the on-rate of DNA binding to dCas9³² or in vivo Cas9 activity^{28, 30–31}. In human cell-based enhanced green fluorescent protein (EGFP) disruption assays, the cleavage activity of SpyCas9 was significantly reduced when double mismatches occur outside the seed region^{28,31}, suggesting a secondary but important role of RDR in reinforcing Cas9 RNP-protospacer interactions. Taken together, the computed buried surface area of the target DNA seems to be a structural indicator of the experimentally observed offtarget sensitivity.

Discussion

Cas9 RNP recognizes protospacers through inherently specific Watson-Crick interactions between the guide RNA and the target strand. Previous work showed, however, that the non-specific interactions imposed by the Cas9 protein on the target also influence specificity. Here, we learned by directed evolution of a thermophilic Cas9, AceCas9, that deleterious mismatches to the wild-type AceCas9 guide RNA may be tolerated by increasing the contact area around the mismatches. We demonstrated that the phosphate lock residues are tunable to lower the sensitivity of Cas9 for mismatches within the seed region. Our result is consistent with the previous finding that the phosphate lock loop promotes pairing between the target strand and the sgRNA⁴⁵ and further reveals the structural principle of how the phosphate lock controls specificity. Potentially, phosphate lock residues could also be tuned to increase the sensitivity of AceCas9. Indeed, mutations of the phosphate lock residues in SpyCas9, KE to GG, rendered SpyCas9 to be sensitive to mismatches at positions⁴⁵.

The fact that mutation of Glu839 to arginine rendered AceCas9 active on T(−1)G protospacer suggests a favorite interaction of the AceCas9 variants for T(−1)G, which could result from a reduced off rate. Both the positive charge and the bulky alkyl chain of Arg839 may fit better than Glu839 to the minor groove of the protospacer helix. Support for this rationalization requires visualization of AceCas9 bound with a substrate DNA.

The effect of Glu839Arg mutation is further enhanced by Glu840Tyr. Though the phosphate lock residues are not strictly conserved (Supplemental figure 2) among Cas9, the position equivalent to AceCas9 Glu840 is invariably a negatively charged residue (Glu or Asp), suggesting a role of electrostatic forces in maintaining specificity at the +1 position. In absence of the negative charge and regardless of the size such as in our selected RY or RA mutant, AceCas9 lost the ability to secure the base pairing. In the known crystal structures of Cas9 bound with a substrate, the Glu or Asp residue is situated in an electrostatically negative environment comprised of ribose oxygen, the non-bridging and the bridging oxygen of the +1 phosphate. The unfavorable interaction instigated by this negative residue could serve to balance other forces such as that from the significantly buried solvent access area acting upon the base pair.

Interaction of the phosphate lock residues with the +1 base pair also impact specificity at other positions within the seed region but not those distal to PAM. We demonstrated that the RY and RA mutants allow AceCas9 to tolerate a mismatch at +4 target position that is poorly recognized by the wild-type AceCas9. We suggest that the alteration of the interaction of the RY or RA mutants with the +1 base pair could increase the overall rate of base annealing between the target strand and the guide RNA, resulting in cleavage of off-targets containing mismatches within the seed region. Detailed kinetics data are required to confirm this hypothesis.

Computationally, we also observed a general correlation between the known specificity data at protospacer positions other than +1 to their buried solvent accessible area in four different Cas9. Buried surface area has been shown to be a strong indicator of binding strength of protein oligomerization⁶¹ and protein-DNA interactions⁶². In the case of protein

oligomerization, buried surface area is a better indicator than chemical composition of the interface in predicting the strength of interaction until it levels off about 2000 Å²⁶¹. Similar trends hold true for the protein-DNA complexes, although protein-DNA interfaces contain significantly more polar and hydrogen bond interactions⁶³. Non-specific protein-DNA contacts such as those between Cas9 and the target DNA, however, have in general reduced hydrogen bond interactions, arguing for an importance of the buried surface area similar to protein-protein interactions⁶². Both previously identified high specificity SpyCas9^{36–37, 45} and our AceCas9 mismatch tolerant variants seem to suggest an importance of buried surface area in specificity. We suggest that specific recognition of the target DNA requires a balance among the sgRNA:DNA base pairing, Cas9 contacts with the target strand, and electrostatic interactions. Highly buried nucleotides depend less on base-pairing with the guide RNA and vice versa.

The measured off-target effect at a given position has been shown to vary among different DNA targets^{28, 31–32}. This variability may be attributable to the different structural environment of different substrates. The computed buried surface area for each Cas9 target was carried out with a specific bound substrate and may, therefore, change slightly when a substrate of different sequence is bound. Alternatively, the buried surface area of the target may be influenced by genome locations or chemical modifications. When combined with kinetic parameters, structural parameters such as the buried surface area can provide more complete picture and allow interpretation of the differences among individual targets and Cas9 enzymes.

Materials and Methods

Plasmid and Library Construction

The plasmid p11-LacY-wtx1 that contains the arabinose-inducible *ccdB*-encoding gene was a generous gift from Dr. David Edgell (University of Western Ontario). The target sequence (ggtaggatggcaagatcctggtatACACC) was placed downstream of the *ccdB* gene between XbaI and HindIII sites to form ccdB-V2. Point mutations in the protospacer (ggtaggatggcaagatcctggtat) were made via Q5 site-directed mutagenesis (NEB) to yield ccdB-N#N where N denotes specific alternations in the protospacer region and # represents the position within the protospacer on the nontarget strand.

The pACYCDuet vector (Novagen 71147) was used to construct pACYCDuet-Ac9-g123 that contains both AceCas9 and its sgRNA-encoding genes, as previously described⁴⁸. To construct the RP-linker AceCas9 library, pACYCDuet-Ac9-g123 was used as the Polymerase Chain Reaction (PCR) template for PCR reactions with synthetic oligos containing randomized codons at residue positions 839–840 as primers. The PCR product was ligated and transformed into DH5α cells. About 3000 colonies were obtained after overnight growth on agar plates containing chloramphenicol and were pooled, grown for 2–5 hours, then collected for extraction of plasmid DNA. The DNA plasmids were sequence verified using Applied Biosystems 3730 Genetic Analyzer with Capillary Electrophoresis before being used for in vivo survival assays. The library represents ~73% of the total possible variants.

In vivo selection through survival assay

The in vivo selection through the survival assay was carried out as previously described (Chen and Zhao, 2005,⁴⁸ with minor adjustments. Selection stains were prepared by transforming *E. coli* BW25141 (gift from D. Edgell) with an appropriate *ccdB*-V2 plasmid, grown in recovery medium containing 0.2% glucose and plated on LB agar plates containing 150 µg/mL ampicillin. Electrocompetent cells of the strain were prepared from a single colony via a standard protocol (Dower et al. 1988). For selection assays, 50 µL of the electrocompetent cells was transformed with 100–200 ng of AceCas9 library or mutant plasmid, and immediately recovered in pre-warmed Super Optimal Broth (SOC) media with shaking at 37 °C for 30 minutes. Then, 0.05 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added followed by incubation at 37 °C with shaking for an additional 60 minutes. For libraries transformation, transformants were concentrated to 200 µL by centrifuging at 3000 ×G for 5 minutes, removing supernatant, and resuspending cell pellet in 200 µL of fresh SOC media. Cells were then plated on LB agar plates containing either chloramphenicol or chloramphenicol and 10 mM arabinose and incubation at 37 °C for 16–20 hours. For specific mutant transformations, transformants were plated directly onto appropriate plates. Colonies were counted manually to estimate survival rates by dividing CFU of the arabinose-containing plate by that of the chloramphenicol only plate after accounting for the dilution factor.

Next Generation Sequencing

Targeted fragments containing phosphate lock library from each pool were amplified via PCR, then indexed Illumina libraries were generated from PCR products (NEBNext Ultra II DNA Library kit for Illumina). Pooled libraries were subjected to 150 cycle single-end sequencing using MiSeq (Illumina). Data was analyzed using Geneious 9.0.5 NGS analysis tools.

Protein Expression and Purification

AceCas9 and mutant variants were expressed and purified as previously described⁴⁸. Briefly, harvested cells were resuspended in lysis buffer and lysed via sonication. Lysate was clarified and loaded to Ni-NTA agarose (Quigen) column. Following Ni-purification, elutant was further purified via a HiTrap Heparin-HP (GE Healthcare) column then on the size exclusion column HiLoad 26/60 Superdex 200 (GE Healthcare). Peak fractions were concentrated and flash frozen in liquid nitrogen and stored at –80 °C.

In vitro transcription and Purification of sgRNA

The reverse complement sequence of sgRNA fused with T7 promoter sequence was annealed in equimolar to a T7 promoter oligo in 10x PCR buffer by incubating in boiling water for 5 minutes, followed by slow cooling. In vitro transcription was initiated by adding 1 µM of annealed dsDNA template into 50 mM Tris-HCl pH 8.0, 40 mM MgCl₂, 10 mM dithiothreitol (DTT), 2 mM spermidine, 5 mM ATP, UTP, CTP, and GTP pH 8.0, 0.1% Triton X-100, and 0.15 mg/mL T7 RNA polymerase. Transcription reaction was incubated at 37 °C overnight, then quenched by adding 50 mM EDTA. The sgRNA was extracted by Phenol:chloroform extraction, and purified via 10% denaturing PAGE.

Plasmid DNA Cleavage Assay

Plasmid cleavage assays were performed as previously described⁴⁸. Briefly, about 6 nM of plasmid DNA was incubated with 500 nM AceCas9:sgRNA in cleavage buffer (20 mM Tris pH 7.5, 150 mM KCl, 2 mM DTT, 10 mM MgCl₂, 5% glycerol) at 50 °C for 10 minutes. Reactions were stopped with 5x gel loading buffer (25 mM Tris pH 7.5, 250 mM EDTA pH 8.0, 1% SDS, 0.05% w/v bromophenol blue, 30% glycerol) and resolved on a 1x TBE 0.5% agarose gel with EZ-Vision In-Gel solution (Amresco). Gels were imaged via a ChemiDoc XRS System (Rio-Rad).

AceCas9 Structural Model

The AceCas9 model was generated using I-TASSER Protein Structure and Function predictor with CjeCas9 (Type II-C) as a model template. The best predicted model has a c score of −0.52 and a predicted RMSD of 10.5 ± 4.6 Å. The proteins predicted to be structurally close to AceCas9 were SauCas9 (5axw) (RMSD = 3.07 Å) and CjeCas9 (5x2g) (RMSD = 3.30 Å). Secondary structures of modeled AceCas9 were compared to those of CjeCas9 and SauCas9 and were found to be largely consistent (Figure S3).

Calculation of Buried Surface Areas

Buried surface areas of each target nucleotide of the four Cas9 whose target-bound crystal structures are available was carried out by using the PDBePISA server (<http://www.ebi.ac.uk/msd-srv/protint/cgi-bin/piserver>) and plotted as a function of nucleotide positions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank M. Tsui and J. Pennington for discussion; B. Miller for advice on in vivo studies; D. Edgell for providing cells and plasmids for in vivo study; H. Moehring and E. Shiel for assistance in preparing and maintaining DNA plasmids; S. Fuller, A. Fisch, H. Pascoe, T. Reynolds for data collection; B. Washburn and C. Pye of the FSU Molecular Cloning Facility for the cloning experiments and discussion. This work was supported by NIH grant R01 GM099604 to H.L.

References

1. Sorek R; Lawrence CM; Wiedenheft B, CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 2013, 82, 237–66. [PubMed: 23495939]
2. Mohanraju P; Makarova KS; Zetsche B; Zhang F; Koonin EV; van der Oost J, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 2016, 353 (6299), aad5147.
3. Koonin EV; Makarova KS; Zhang F, Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology* 2017, 37, 67–78. [PubMed: 28605718]
4. Shmakov S; Smargon A; Scott D; Cox D; Pyzocha N; Yan W; Abudayyeh OO; Gootenberg JS; Makarova KS; Wolf YI; Severinov K; Zhang F; Koonin EV, Diversity and evolution of class 2 CRISPR-Cas systems. *Nature reviews. Microbiologssssy* 2017, 15 (3), 169–182.
5. Deltcheva E C. K. S. C. M. G. K. C. Y. P. Z. A. E. M. R. V. J. C. E., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011, 471 (7340), 602–607. [PubMed: 21455174]

6. Tsui TK; Li H, Structure Principles of CRISPR-Cas Surveillance and Effector Complexes. *Annu Rev Biophys* 2015.
7. Garcia-Doval C; Jinek M, Molecular architectures and mechanisms of Class 2 CRISPR-associated nucleases. *Curr Opin Struct Biol* 2017, 47, 157–166. [PubMed: 29107822]
8. Lewis KM; Ke A, Building the Class 2 CRISPR-Cas Arsenal. *Mol Cell* 2017, 65 (3), 377–379. [PubMed: 28157502]
9. Nunez JK; Harrington LB; Doudna JA, Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chem Biol* 2016.
10. Barrangou R; Marraffini LA, CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* 2014, 54 (2), 234–44. [PubMed: 24766887]
11. Jinek M C. K. F. I. H. M. D. J. A. C. E., A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. 2012; Vol. 337, pp 816–821.
12. Hou Z; Zhang Y; Propson NE; Howden SE; Chu LF; Sontheimer EJ; Thomson JA, Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America* 2013, 110 (39), 15644–9. [PubMed: 23940360]
13. Esvelt KM; Mali P; Braff JL; Moosburner M; Yaung SJ; Church GM, Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods* 2013, 10 (11), 1116–21. [PubMed: 24076762]
14. Ran FA; Cong L; Yan WX; Scott DA; Gootenberg JS; Kriz AJ; Zetsche B; Shalem O; Wu X; Makarova KS; Koonin EV; Sharp PA; Zhang F, In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 2015, 520 (7546), 186–91. [PubMed: 25830891]
15. Heidenreich M; Zhang F, Applications of CRISPR-Cas systems in neuroscience. *Nat Rev Neurosci* 2016, 17 (1), 36–44. [PubMed: 26656253]
16. Wright AV; Nunez JK; Doudna JA, Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* 2016, 164 (1–2), 29–44. [PubMed: 26771484]
17. Strong A; Musunuru K, Genome editing in cardiovascular diseases. *Nat Rev Cardiol* 2017, 14 (1), 11–20. [PubMed: 27609628]
18. Han Y; Slivano OJ; Christie CK; Cheng AW; Miano JM, CRISPR-Cas9 genome editing of a single regulatory element nearly abolishes target gene expression in mice--brief report. *Arterioscler Thromb Vasc Biol* 2015, 35 (2), 312–5. [PubMed: 25538209]
19. Ding Q; Strong A; Patel KM; Ng SL; Gosis BS; Regan SN; Cowan CA; Rader DJ; Musunuru K, Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ Res* 2014, 115 (5), 488–92. [PubMed: 24916110]
20. Long C; McAnally JR; Shelton JM; Mireault AA; Bassel-Duby R; Olson EN, Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* 2014, 345 (6201), 1184–1188. [PubMed: 25123483]
21. Wang Y; Du Y; Shen B; Zhou X; Li J; Liu Y; Wang J; Zhou J; Hu B; Kang N; Gao J; Yu L; Huang X; Wei H, Efficient generation of gene-modified pigs via injection of zygote with Cas9/sgRNA. *Sci Rep* 2015, 5, 8256. [PubMed: 25653176]
22. Sander JD; Joung JK, CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 2014, 32 (4), 347–55. [PubMed: 24584096]
23. Doudna JA; Charpentier E, Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014, 346 (6213), 1258096.
24. Wang H; Yang H; Shivalila CS; Dawlaty MM; Cheng AW; Zhang F; Jaenisch R, One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 2013, 153 (4), 910–8. [PubMed: 23643243]
25. Shen B; Zhang J; Wu H; Wang J; Ma K; Li Z; Zhang X; Zhang P; Huang X, Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* 2013, 23 (5), 720–3. [PubMed: 23545779]
26. Niu Y; Shen B; Cui Y; Chen Y; Wang J; Wang L; Kang Y; Zhao X; Si W; Li W; Xiang AP; Zhou J; Guo X; Bi Y; Si C; Hu B; Dong G; Wang H; Zhou Z; Li T; Tan T; Pu X; Wang F; Ji S; Zhou Q; Huang X; Ji W; Sha J, Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 2014, 156 (4), 836–43. [PubMed: 24486104]

27. DiCarlo JE; Norville JE; Mali P; Rios X; Aach J; Church GM, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 2013, 41 (7), 4336–43. [PubMed: 23460208]
28. Fu Y; Foden JA; Khayter C; Maeder ML; Reyon D; Joung JK; Sander JD, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013, 31 (9), 822–6. [PubMed: 23792628]
29. Hsu PD S. D. A. W. J. A. R. F. A. K. S. A. V. L. Y. F. E. J. , DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* 2013, 31 (9), 827–832.
30. Pattanayak V; Lin S; Guilinger JP; Ma E; Doudna JA; Liu DR, High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 2013, 31 (9), 839–43. [PubMed: 23934178]
31. Anderson EM; Haupt A; Schiel JA; Chou E; Machado HB; Strezoska Z; Lenger S; McClelland S; Birmingham A; Vermeulen A; Smith A, Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *J Biotechnol* 2015, 211, 56–65. [PubMed: 26189696]
32. Boyle EA; Andreasson JOL; Chircus LM; Sternberg SH; Wu MJ; Guegler CK; Doudna JA; Greenleaf WJ, High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci US A* 2017, 114 (21), 5461–5466.
33. Fu BX; Hansen LL; Artiles KL; Nonet ML; Fire AZ, Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res* 2014, 42(22), 13778–87. [PubMed: 25399416]
34. Cong L; Ran FA; Cox D; Lin S; Barretto R; Habib N; Hsu PD; Wu X; Jiang W; Marraffini LA; Zhang F, Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, 339 (6121), 819–23. [PubMed: 23287718]
35. Tsai SQ; Zheng Z; Nguyen NT; Liebers M; Topkar VV; Thapar V; Wyvekens N; Khayter C; Iafrate AJ; Le LP; Aryee MJ; Joung JK, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* 2015, 33 (2), 187–197. [PubMed: 25513782]
36. Kleinstiver BP; Pattanayak V; Prew MS; Tsai SQ; Nguyen NT; Zheng Z; Joung JK, High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016, 529 (7587), 490–5. [PubMed: 26735016]
37. Slaymaker IM; Gao L; Zetsche B; Scott DA; Yan WX; Zhang F, Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016, 351 (6268), 84–8. [PubMed: 26628643]
38. Raper AT; Stephenson AA; Suo Z, Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J Am Chem Soc* 2018, 140 (8), 2971–2984. [PubMed: 29442507]
39. Gong S; Yu HH; Johnson KA; Taylor DW, DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep* 2018, 22 (2), 359–371. [PubMed: 29320733]
40. Bisaria N; Jarmoskaite I; Herschlag D, Lessons from Enzyme Kinetics Reveal Specificity Principles for RNA-Guided Nucleases in RNA Interference and CRISPR-Based Genome Editing. *Cell Syst* 2017, 4 (1), 21–29. [PubMed: 28125791]
41. Klein M; Eslami-Mossallam B; Arroyo DG; Depken M, Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. *Cell reports* 2018, 22 (6), 1413–1423. [PubMed: 29425498]
42. Ran FA H. P. D. L. C.-Y. G. J. S. K. S. T. A. E. S. D. A. I. A., Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 2013, 154 (6), 1380–1389. [PubMed: 23992846]
43. Mali P Y. L. E. K. M. A. J. G. M. D. J. E. N. J. E. C. G., RNA-Guided Human Genome Engineering via Cas9. *Science* 2013, 339 (6121), 823–826. [PubMed: 23287722]
44. Guilinger JP; Thompson DB; Liu DR, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 2014, 32 (6), 577–582. [PubMed: 24770324]
45. Anders C; Niewoehner O; Duerst A; Jinek M, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014, 513 (7519), 569–73. [PubMed: 25079318]
46. Chen JS; Dagdas YS; Kleinstiver BP; Welch MM; Sousa AA; Harrington LB; Sternberg SH; Joung JK; Yildiz A; Doudna JA, Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017, 550 (7676), 407–410. [PubMed: 28931002]
47. Chen JS D. Y. S. K. B. P. W. M. M. S. A. A. H. L. B. S. S. H. , Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017, 550 (7676), 407–410. [PubMed: 28931002]

48. Tsui TKM; Hand TH; Duboy EC; Li H, The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9. *ACS Synth Biol* 2017, 6 (6), 1103–1113. [PubMed: 28277645]
49. Mir A; Edraki A; Lee J; Sontheimer EJ, Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* 2018, 13 (2), 357–365. [PubMed: 29202216]
50. Chen Z; Zhao H, A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res* 2005, 33 (18), e154. [PubMed: 16214805]
51. Roy A; Kucukural A; Zhang Y, I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010, 5 (4), 725–38. [PubMed: 20360767]
52. Yang J; Yan R; Roy A; Xu D; Poisson J; Zhang Y, The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 2015, 12 (1), 7–8. [PubMed: 25549265]
53. Zhang Y, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008, 9, 40. [PubMed: 18215316]
54. Yamada M; Watanabe Y; Gootenberg JS; Hirano H; Ran FA; Nakane T; Ishitani R; Zhang F; Nishimasu H; Nureki O, Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Mol Cell* 2017, 65 (6), 1109–1121 e3. [PubMed: 28306506]
55. Nishimasu H; Cong L; Yan WX; Ran FA; Zetsche B; Li Y; Kurabayashi A; Ishitani R; Zhang F; Nureki O, Crystal Structure of *Staphylococcus aureus* Cas9. *Cell* 2015, 162 (5), 1113–26. [PubMed: 26317473]
56. Nishimasu H; Ran FA; Hsu PD; Konermann S; Shehata SI; Dohmae N; Ishitani R; Zhang F; Nureki O, Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* 2014.
57. Hirano H; Gootenberg JS; Horii T; Abudayyeh OO; Kimura M; Hsu PD; Nakane T; Ishitani R; Hatada I; Zhang F; Nishimasu H; Nureki O, Structure and Engineering of *Francisella novicida* Cas9. *Cell* 2016, 164 (5), 950–61. [PubMed: 26875867]
58. Cho SW; Kim S; Kim Y; Kweon J; Kim HS; Bae S; Kim JS, Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 2014, 24 (1), 132–41. [PubMed: 24253446]
59. Hsu PD; Scott DA; Weinstein JA; Ran FA; Konermann S; Agarwala V; Li Y; Fine EJ; Wu X; Shalem O; Cradick TJ; Marraffini LA; Bao G; Zhang F, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* 2013, 31 (9), 827–32.
60. Jinek M; Jiang F; Taylor DW; Sternberg SH; Kaya E; Ma E; Anders C; Hauer M; Zhou K; Lin S; Kaplan M; Iavarone AT; Charpentier E; Nogales E; Doudna JA, Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 2014, 343 (6176), 1247997.
61. Chen J; Sawyer N; Regan L, Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* 2013, 22 (4), 510–5. [PubMed: 23389845]
62. Jones S; van Heyningen P; Berman HM; Thornton JM, Protein-DNA interactions: A structural analysis. *Journal of molecular biology* 1999, 287 (5), 877–96. [PubMed: 10222198]
63. Ramos RM; Fernandes LF; Moreira IS, Extending the applicability of the O-ring theory to protein-DNA complexes. *Comput Biol Chem* 2013, 44, 31–9. [PubMed: 23537566]

Figure 1: Structural Features of AceCas9 and the Guide RNA Constructs Used in This Study.

(A) Cartoon of AceCas9 showing different domains and location of mismatches on the nontarget strand of the substrate plasmid. Nucleotide positions for both nontarget and the target strands are labeled. (B) A structure model of AceCas9 highlighting the protein domains and features around the target DNA. Abbreviations: REC, recognition domain, gray; HNH, HNH catalytic domain, pink; RuvC, RuvC catalytic domain, green; PID, PAM interacting domain, cyan; target DNA shown in black; sgRNA shown in red.

A. **B.**

Figure 2. Sensitivity of AceCas9 to Protospacer Mismatches Revealed by a Bacterial Survival Assay.

(A) Cartoon illustration of the in-cell survival assay. Electrocompetent *E. coli* harboring an arabinose inducible *ccdB* plasmid containing AceCas9 PAM and an either matching or mismatched protospacer is transformed with a dual expression vector encoding Cas9 and its sgRNA. Expression of functional AceCas9 and its sgRNA results in the cleavage of *ccdB* plasmid, and thus cell survival in the presence of arabinose (Ara+). Incomplete cleavage of the *ccdB* plasmid due to mismatches between the sgRNA and target DNA leads to cell death in the presence of arabinose. (B) Images of Ara+ and Ara- plates are shown for cells

harbouring different substrate plasmids with the calculated survival rate shown as the bar diagram directly below.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Next Generation Sequencing Results of the Survival Mutants.

(A) Cartoon illustration of iterative rounds of survival with AceCas9 library. (B) NGS results for each of the 3 rounds of survival, as well as the initial library. (C) A close-up view of AceCas9 structure model near the +1 target position indicating possible interactions with Glu839 and Glu840. (D) Results of iterative rounds of directed evolution of Glu839/Glu840. Left, an example of plates after the 3rd round of survival is shown. Right, amino acid sequences among the survival mutants are listed along those of the wild-type AceCas9.

Figure 4. Activity Verification of Selected Variants.

(A) Plasmids encoding the selected AceCas9 mutants were transformed into the ccdB-T(-1)G harbouring cells to confirm enhanced survival. Images of plates under Ara+ and Ara- are shown for each mutant with the calculated survival rate shown as the bar diagram directly below. (B) Comparison of in vitro DNA cleavage activities on the wild-type and the T(-1)G plasmids among the purified AceCas9. Molar excess AceCas9 (wild-type and mutants) were incubated with plasmids for 10 minutes at 50 °C before being separated and visualized on agarose gels.

Figure 5. Sensitivity of T(-1)G AceCas9 Mutants to Other Mismatches.

(A) In cell and in vitro activity assays with the wild-type AceCas9, the RY and the RA mutants on DNA substrates with mismatches at -4 and -20 to -24 (Fk-AA and Fk-AAGC) nontarget positions. Images of plates under Ara+ and Ara- are shown for each variant with the calculated survival rate shown as the bar diagram directly below. (B) Comparison of in vitro DNA cleavage activities on the wild-type and the G(-4)A, Fk-AA, and Fk-AAGC plasmids among the purified AceCas9. Molar excess AceCas9 (wild-type and variants) were incubated with plasmids for 10 minutes at 50 °C before being separated and visualized on agarose gels.

Figure 6. Computed Buried Surface Area of the Target DNA Bound with Four Cas9.

Computed buried surface areas for the target DNA strand by the four known Cas9s are displayed for each nucleotide position. The Protein Data Bank codes for each of the known Cas9 structures are shown next to the abbreviated names of Cas9. The buried surface areas were computed by use of the PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver).