IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No.	:	15/632,067	Conf. No.: 1196
First Named Inventor	:	Feng Zhang	
Filed	:	June 23, 2017	
TC/AU	:	1633	
Examiner	:	Maria Gomez Leavitt	
Customer No.	:	198696	
Docket No.	:	BROD-3260US	
Title	:	RNA-TARGETING SYSTEM	

AMENDMENT AND RESPONSE

FILED VIA EFS-WEB

Commissioner:

In response to the Final Office Action mailed December 19, 2023 ("Office Action"), please enter and consider the following amendments and remarks.

Certificate of EFS-WEB Transmission pursuant to 37 C.F.R. § 1.8

I hereby certify that this correspondence is being transmitted via the U.S. Patent and Trademark Office electronic filing system (EFS-Web) to the USPTO, on May 17, 2024

/Christopher Southgate/

Christopher D. Southgate, Ph.D., Esq.

Reg. No. 54,875

AMENDMENT

In the Claims

Please amend the claims as follows. This listing of claims will replace all prior versions and listings of claims in this application.

1. (Currently Amended) A non-naturally occurring or engineered vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to a first polynucleotide sequence encoding an RNA-targeting guide RNA (Rt-gRNA), wherein said Rt-gRNA is designed to hybridize with a target RNA, wherein said Rt-gRNA comprises:

(i) a small CRISPR/Cas system associated RNA (scaRNA) sequence, and

(ii) a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence, wherein said scaRNA and said tracrRNA are designed to at least partially hybridize to each other, and

b) a second regulatory element operably linked to a second polynucleotide sequence encoding

(i) a *Francisella novicida* Cas protein (FnCas) or ortholog, homolog, or fragment thereof comprising a Rec1 domain, Rec2 domain, HNH domain, and PAM Interacting (PI) domain, and

(ii) one or more heterologous functional domains positioned at one or more sites ~~on the~~ in ~~FnCas9~~ FnCas that are analogous to [the] amino acid positions 94–179 and 308–713 ~~534-676~~ corresponding to ~~of the~~ a *Streptococcus pyogenes* Cas9 (SpCas9) Rec1 domain,

wherein the Rt-gRNA is capable of forming an RNA-targeting complex with the FnCas and of directing sequence-specific binding of the RNA-targeting complex to the target RNA,

wherein each of the one or more heterologous functional domains consists of methylase activity, demethylase activity, translation activation activity, translation repression activity, histone modification activity, RNA cleavage activity, or DNA cleavage activity, and

wherein components (a) and (b) are located on the same or different vectors of the system and wherein the Rt-gRNA and said FnCas protein do not naturally occur together.

2. (Canceled)

3. (Previously Presented) The vector system according to claim 1, wherein said FnCas protein is a type II Cas protein.

4-8. (Canceled)

9. (Original) The vector system according to claim 1, wherein said scaRNA sequence is fused to said tracrRNA sequence.

10. (Previously Presented) The vector system according to claim 1, wherein said FnCas protein is codon optimized for expression in a eukaryotic cell.

11. (Previously Presented) The vector system according to claim 1, wherein said one or more vectors are viral vectors, or wherein said one or more vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

12. (Canceled)

13. (Previously Presented) The vector system according to claim 1, wherein said RNA-targeting guide RNA is designed to direct said FnCas protein to said target RNA.

14. (Withdrawn) A vector comprising a regulatory element operably linked to a polynucleotide sequence encoding an RNA-targeting Cas protein, preferably FnCas9 or an FnCas9 homolog or ortholog.

15. (Withdrawn) A vector comprising a regulatory element operably linked to a polynucleotide sequence encoding an RNA-targeting guide RNA, wherein said RNA-targeting guide RNA is capable of hybridizing with a target RNA, wherein said RNA-targeting guide RNA comprises:

(i) a small CRISPR/Cas system associated RNA (scaRNA) sequence, and

(ii) a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence, wherein said scaRNA and said tracrRNA are capable of at least partially hybridizing.

16. (Previously Presented) A composition comprising a vector system according to claim 1.

17. (Withdrawn) A non-naturally occurring or engineered RNA-targeting system comprising an RNA-targeting Cas protein and at least one RNA-targeting guide RNA, wherein said RNA-targeting guide RNA is capable of binding to a target RNA in a eukaryotic cell, wherein said RNA-targeting guide RNA comprises:

(i) a small CRISPR/Cas system associated RNA (scaRNA) sequence, and

(ii) a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence, wherein said scaRNA and said tracrRNA are capable of at least partially hybridizing.

18. (Withdrawn) The system according to claim 17, wherein said RNA-targeting Cas protein is FnCas9 or an FnCas9 homolog or ortholog.

19. (Withdrawn) A method for modulating synthesis of a protein comprising introducing into a cell containing an RNA molecule encoding said protein, a vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to a polynucleotide sequence encoding an RNA-targeting guide RNA, wherein said RNA-targeting guide RNA is capable of hybridizing with said target RNA, wherein said RNA-targeting guide RNA comprises:

(i) a small CRISPR/Cas system associated RNA (scaRNA) sequence, and

(ii) a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence, wherein said scaRNA and said tracrRNA are capable of at least partially hybridizing, and

b) a second regulatory element operably linked to a polynucleotide sequence encoding an RNA-targeting Cas protein, wherein components (a) and (b) are located on the same or different vectors of the system, whereby synthesis of said protein is modulated.

20. (Withdrawn) The method according to claim 19, wherein the synthesis of said protein is suppressed; or the synthesis of said protein is modulated by editing of said RNA molecule encoding said protein or splicing of said RNA molecule encoding said protein.

21. (Withdrawn) The method according to claim 20, wherein the synthesis of said protein is suppressed by knock-down of said RNA molecule encoding said protein.

22-23. (Canceled)

24. (Withdrawn) A method for targeting RNA comprising introducing into a cell containing a target RNA a non-naturally occurring or engineered vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to a polynucleotide sequence encoding an RNA-targeting guide RNA, wherein said RNA-targeting guide RNA is capable of hybridizing with said target RNA, wherein said RNA-targeting guide RNA comprises:

(i) a small CRISPR/Cas system associated RNA (scaRNA) sequence, and

(ii) a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence, wherein said scaRNA and said tracrRNA are capable of at least partially hybridizing, and

b) a second regulatory element operably linked to a polynucleotide sequence encoding an RNA-targeting Cas protein, wherein components (a) and (b) are located on the same or different vectors of the system, whereby said RNA-targeting guide RNA directs said RNA-targeting Cas protein to said target RNA.

25. (Withdrawn) The method according to claim 19, wherein said RNA-targeting Cas protein is a type II Cas protein.

26. (Withdrawn) The method according to claim 25, wherein said RNA-targeting Cas protein is a Cas9 protein, is from *Francisella novicida*, is an FnCas9 homolog showing at least 80% sequence homology with wild type FnCas9, or is an FnCas9 ortholog.

27-29. (Canceled)

30. (Withdrawn) The method according to claim 19, wherein said RNA-targeting Cas protein is codon optimized for expression in a eukaryotic cell.

31. (Withdrawn) The method according to claim 19, wherein said one or more vectors are viral vectors, or wherein said one or more vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

32. (Canceled)

33. (Withdrawn) The method according to claim 19, wherein the cell is a eukaryotic cell.

34. (Withdrawn) The method according to claim 30, wherein the eukaryotic cell is a mammalian or human cell.

35. (Previously Presented) The composition according to claim 16, wherein one or more amino acid residues of the FnCas protein, other than those amino acid positions at which the one or more heterologous functional domains are positioned, are modified.

36. (Previously Presented) The composition according to claim 35, wherein the modification comprises mutation of one or more, or two or more, or three or more amino acid residues of the FnCas protein.

37. (Previously Presented) The composition according to claim 36, wherein the one or more, or two or more, or three or more mutations are in one or more catalytically active domains of the FnCas protein.

38. (Previously Presented) The composition according to claim 37, wherein the FnCas protein has reduced or abolished nuclease activity compared with an FnCas protein lacking said mutation(s).

39. (Previously Presented) The composition according to claim 38, wherein the one or more, or two or more mutations are in a catalytically active domain of the FnCas protein comprising a RuvCI, RuvCII or RuvCIII domain.

40. (Canceled)

41. (Previously Presented) The vector system according to claim 1, wherein the FnCas further comprises one or more nuclear localization signals.

42-43. (Canceled)

44. (Withdrawn) The vector system according to claim 41, wherein the transcriptional activation domain comprises VP64.

45. (Canceled)

46. (Withdrawn) The vector system according to claim 41, wherein the transcriptional repression domain comprises a KRAB domain or a SID domain.

47. (Canceled)

48. (Withdrawn) The vector system according to claim 41, wherein a nuclease domain comprises Fok1.

49. (Previously Presented) The vector system according to claim 1, wherein the one or more functional domains have one or more of the following activities: deaminase activity, methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity and nucleic acid binding activity.

50-56. (Canceled)

57. (Previously Presented) The composition according to claim 16, wherein the tracrRNA or the scaRNA comprises one or more protein-binding RNA aptamers.

58. (Canceled)

59. (Previously Presented) The composition according to claim 57, wherein the one or more aptamers is designed to bind a bacteriophage coat protein.

60. (Previously Presented) The composition according to claim 59, wherein the bacteriophage coat protein is selected from the group consisting of Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

61. (Canceled)

62. (Previously Presented) The composition according to claim 16, wherein the tracrRNA is 30 or more, 40 or more or 50 or more nucleotides in length.

63. (Previously Presented) The composition according to claim 16, wherein the FnCas protein is modified to comprise a PAM sequence specificity which is different from the PAM sequence specificity of the FnCas protein without said further modification.

64. (Previously Presented) The composition according to claim 63, wherein said modification comprises the introduction of one or more amino acid mutations into the FnCas protein, or by truncation of the FnCas protein, or by deletion and/or insertion of specific amino acids or amino acid sequences into the FnCas protein.

65. (Previously Presented) The composition according to claim 16, wherein the vector comprises a plurality of polynucleotide sequences encoding multiple Rt-gRNAs, wherein each Rt-gRNAs is specific for a different target RNA whereby there is multiplexing.

66. (Previously Presented) A host cell or cell line comprising the composition according to claim 16, or progeny thereof.

67. (Previously Presented) The host cell or cell line or progeny thereof according to claim 66, which is *ex vivo* or *in vitro*, a stem cell or a stem cell line, or a plant cell or a plant cell line.

68. (Canceled)

69. (Previously Presented) The vector system according to claim 1, wherein the one or more functional domains are associated with an adaptor protein.

70. (Previously Presented) The vector system according to claim 1, wherein the FnCas protein comprises two or more functional domains.

71. (Canceled)

72. (Currently amended) The vector system according to claim 1, wherein the FnCas protein or ortholog, homolog, or fragment thereof, ~~further~~ comprises one or more heterologous functional domains positioned at one or more sites in FnCas that are analogous to amino acid positions 180-307 of a *Streptococcus pyogenes* Cas9 (SpCas9) ~~at one or more amino acid positions 175-306 of the FnCas~~ Rec2 domain.

73. (New) The vector system according to claim 1, wherein the FnCas protein or ortholog, homolog, or fragment thereof, further comprises one or more heterologous functional domains positioned at one or more sites in FnCas that are analogous to amino acid positions 534-676 of a *Streptococcus pyogenes* Cas9 (SpCas9) Rec1 domain.

REMARKS

Status of the Claims

Claims 1, 3, 9–11, 13–21, 24–26, 30–31, 33–39, 41, 44, 46, 48–49, 57, 59–60, 62–67, 69–70, and 72–73 are pending, with claims 1, 14–15, 17, 19, and 24 being independent. Claims 1 and 72 are amended. Claims 2, 4–8, 12, 22–23, 27–29, 32, 40, 42–43, 45, 47, 50–56, 58, 61, 68, and 71 are canceled. Claim 73 is new. Claims 14–15, 17–21, 24–26, 30–31, 33–34, 44, 46, and 48 are withdrawn. All claim amendments are made without prejudice or disclaimer.

No new matter is added.

Priority

The Applicant respectfully disagrees that the subject matter of claim 1 is not supported by provisional application 62/098,059, filed on December 30, 2014, as discussed in the response to the claim rejections under 35 U.S.C. § 112(a) below. Further to the arguments presented below, the Applicant submits that the earliest effective filing date of all the pending claims is December 23, 2014, i.e., the filing date of Provisional Patent Application 62/096,324.

Claim Rejection Under 35 U.S.C. § 112(b) – Indefiniteness

Claims 1, 3, 9–11, 13, 16, 35–39, 41, 49, 57, 59, 60, 62–67, 69-70 stand rejected and claim 72 is newly rejected under 35 U.S.C. § 112(b) as allegedly being indefinite for failing to particularly point out and distinctly claim the subject matter regarded as the invention. Specifically, the Office has found there are no proper antecedent bases for the recitation of "the FnCas9" in the cited claims. Office Action, at p. 9. The Office also finds that the recitation of "one or more sites on the FnCas9 that are analogous to the amino acid positions 534–676 of the *Streptococcus pyogenes* Cas9 (SpCas9) Rec1 domain" is indefinite. *Id.* at p. 10. The Office finds that the present application does not provide an alignment of the amino acid sequences of SpCas9 Rec1 and FnCas9 to locate the sites in FnCas9 that are analogous to *the amino acid positions 534–676 SpCas9 Rec1*." *Id.* (emphasis in original).

The Applicant respectfully disagrees with the Examiner for the following reasons.

The absence of antecedent basis has been addressed by the amendment of the claims presented herein. Regarding the remaining indefiniteness issues, as noted in MPEP § 2173.02, the disclosure, the prior art, and claim interpretation are all factors in determining the definiteness of

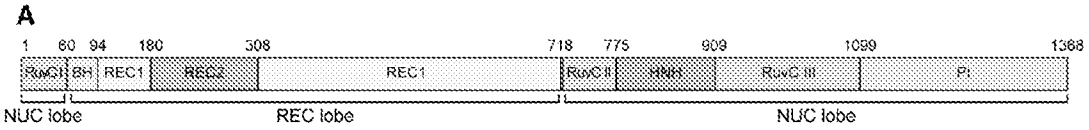
presented claims. The test for definiteness is whether those of ordinary skill in the relevant art would understand what is claimed without an explicit alignment of FnCas9 and SpCas9 Rec1.[1]

A structural comparison of the REC1 domain[2] of *Streptococcus pyogenes* Cas9 (SpCas9) based on crystallographic data with the Cas9 REC1 domains of *Streptococcus mutans*; *Streptococcus thermophilus* CRISPR-3; *Streptococcus thermophilus* CRISPR-1; *Campylobacter jejuni*; *Neisseria meningitidis* was disclosed in Nishimasu et al. (2014)[3] that published on **February 13, 2014**, i.e., before Applicant's earliest filing date of **December 23, 2014**.[4]

Specifically, Nishimasu (2014)[3] states on page 938:

> The crystal structure revealed that Cas9 consists of two lobes, a recognition (REC) lobe and a nuclease (NUC) lobe (Figures 1A–D). The REC lobe can be divided into three regions, a long α-helix referred to as the Bridge helix (residues 60–93), the ***REC1 (residues 94–179 and 308–713) domain***, and the REC2 (residues 180–307) domain (Figures 1A–D). (Emphasis added)

A diagram of SpCas9's domain structure published by Nishimasu (2014) (FIG. 1A) is shown below.



An amino acid sequence comparison of Cas9 Orthologs in Families II-A and II-C can be found in Appendix IA (Nishimasu (2014), Suppl. FIG. S4). Residues 94–179 and 308–713 of the REC1 domain are highlighted (see black rectangles, Appendix I), in addition to residues 534-676, which are referenced in pending claim 1 of this application and reproduced below.

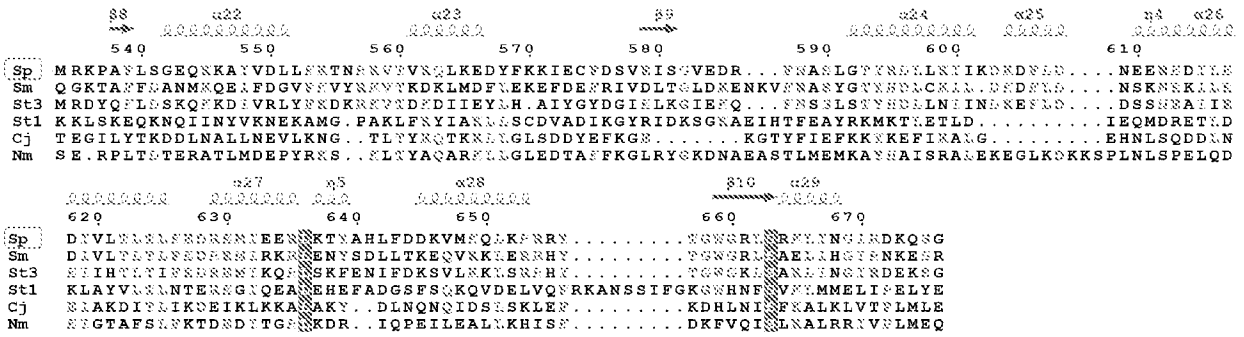
---

[1] ". . .[t]he specification need not disclose what is well-known to those skilled in the art and preferably omits that which is well-known to those skilled and ***already available to the public***." MPEP § 2164.05(a) (emphasis added) (*citing In re Buchner,* 929 F.2d 660, 661, (Fed. Cir. 1991).

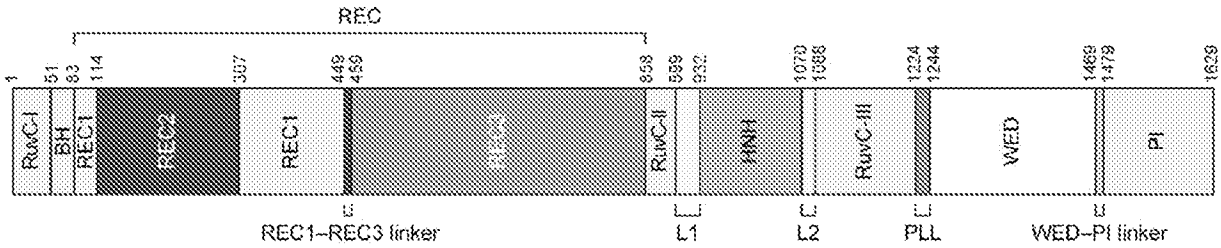
[2] Supplemental Figure S4 of Nishimasu et al. (2014), reproduced in part in Appendix I, shows a sequence alignment of Cas9 orthologs in families II-A and II-C.

[3] Nishimasu et al. (2014) Cell. 156(5):935-49.

[4] filing date of Provisional Patent Application No. 62/096,324



Hirano et al. (2016)[5], published on February 25, 2016, reported on the structure of *Francisella novicida* Cas9. A schematic representation of FnCas9's domain organization of is shown below (Hirano, FIG. 2A).



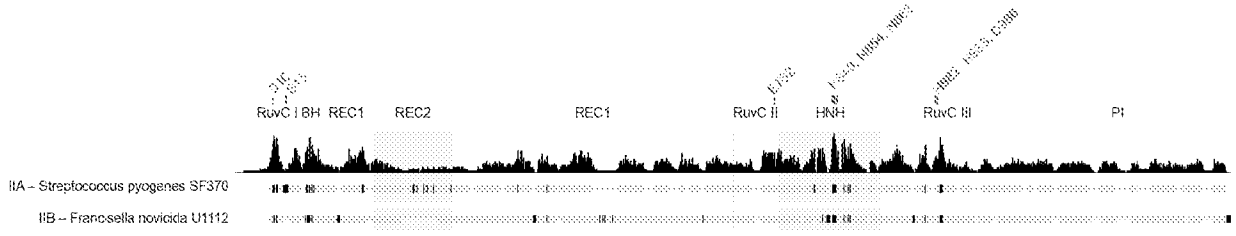
At page 5 of the Office Action, the Examiner states:

> However, there is not an amino acid sequence alignment provided for Francisella novicida Cas and CRISPR-Cas9 (*S. pyogenes*) to locate the analogous position of residues 534- 676 so one of ordinary skill in the art could appreciate which are the analogous sites on the FnCas9 and SpCas9 Recl domain. ***In fact, it was previously noted that location of amino acid positions 534-676 of the SpCas9 does not correspond to the REC1 domain according to post-filing art of Hirano et al. (Cell 164, 950-961, February 25, 2016; of record).*** Hirano et al., teaches the amino acid sequence *Francisella novicida* Cas9 of 1629 amino acids, wherein the REC2 domain is inserted into the REC1 domain, and the REC1 and REC3 domains are connected by a linker loop. (Emphasis added)

Besides the difference in the names, Hirano (2016) fails to indicate how "REC3", "REC2" or "REC1" of FnCas9 relate to corresponding REC1 and REC2 domains of SpCas9. Just because the name is similar does not necessarily mean FnCas9 REC1 is equivalent to SpCas9 REC1 or that FnCas9 REC3 is distinct from SpCas9 REC1. In the absence of evidence to the contrary, the designation of "REC3" appears to be entirely arbitrary. In apparent corroboration of this assessment, Supplemental Figure S5 of the 2014 Nishimasu article, reproduced in part below, shows a **sequence alignment** of Cas9 orthologs in families II-A, II-B and II-C, including

[5] Hirano et el. (2016) Cell 164(5): 950–961

Francisella novicida Cas protein (FnCas) and SpCas9 (see Appendix IB).[6] Applicant notes those sequences designated as REC3 by Hirano (2016) are labeled as REC1 by Nishimasu (2014) suggesting the designation of "REC3" by Hirano is meant to emphasize a difference in FnCas9's protein structure as compared to SpCas9.



As of Applicant's filing, the amino acid sequence of SpCas9 and FnCas9 were publicly available, for example, in the database at the UniProtKB/Swiss-Prot Consortium.[7]

The published amino acid sequence of FnCas9 (CAS9_FRATN Accession No: A0Q5Y3) as of **November 26, 2014**, is shown below:

>sp|A0Q5Y3|Release 2014_11/2014_11|26-Nov-2014
MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVYELSKDSYTLLMNNRTARRHQRRGIDRKQLVKRLFKL
IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
SEIYNKLMQKILEFKLMKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYYHH
DKYNIQEFLKRHATINDRILDTLLTDDLDIWNFNFEKFDFDKNEEKLQNQEDKDHIQAHLHHFVFAVNKIKSEMASG
GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLIGNLSNLELKPLRKYFNDKIHAKADHWDE
QKFTETYCHWILGEWRVGVKDQDKKDGAKYSYKDLCNELKQKVTKAGLVDFLLELDPCRTIPPYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQQIASGQRDYKDLDARILQ
FIFDRVKASDELLLNEIYFQAKKLKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLHKYNNTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEG**SEDKKGNYKHGLAYELGVL**
LFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
QRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKSLKDRRKKAL
ERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLICVTRGDNKNKGNR
IFCLRDLADNYKLKQFETTDDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIR
AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDK
PQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDIFSQIKITDNEFSDKKLVRKKAIEG
FNTHRQMTRDGIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFVDKPISIDIQISTL
EELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEFLRSLAYRSERVKIKSIDDVKQVLDKD
SNFIIGKITLPFKKEWQRLYREWQNTTIKDDYEFLKSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVKRKTWDNNF
IYQILNDSDSRADGTKPFIPAFDISKNEIVEAIIDSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLR

[6] Thirty-five Cas9 orthologs from families IIA, IIB and IIC were aligned (BLOSUM62) and clustered (Jukes-Cantor model Neighbor-Joining method, with S. pyogenes Cas9 as the outgroup). Bars on the top show amino acid conservation. In each line, the black bars show residues with at least 75% consensus, and the gray bars show non-conserved residues.

[7] Breuza L, Poux S, Estreicher A, Famiglietti ML, Magrane M, Tognolli M, Bridge A, Baratin D, Redaschi N; UniProt Consortium. The UniProtKB guide to the human proteome. Database (Oxford). 2016 Feb 20;2016:bav120. doi: 10.1093/database/bav120. PMID: 26896845; PMCID: PMC4761109.

DIGIATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEMLGM
KLAGIYNETSNN(Emphasis added)[8]

The published amino acid sequence of SpCas9 (CAS9_STRP1 Accession No: Q99ZW2) as of **November 26, 2014**, is shown below:

>sp|Q99ZW2|Release 2014_11/2014_11|26-Nov-2014
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG**MRKPAF**
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (Emphasis added)[9]

With the availability of both the SpCas9 and FnCas9 amino acid sequences prior to Applicant's earliest effective filing date of December 23, 2014, it would have been routine for a person of ordinary skill in the art to perform a sequence alignment using publicly available software programs to identify amino acids in the FnCas9 that are analogous to amino acid positions 534-676 of SpCas9 within the REC1 domain.

For example, Panda and Ray published in 2022,[10] a pairwise sequence alignment of SpCas9 and FnCas9 sequences using EMBOSS, which implements the Needleman-Wunsch Algorithm to find the optimum alignment (including gaps) of two sequences along their entire length (see Appendix II, Panda and Ray, Suppl. file 2). The EMBOSS software has been publicly available since at least the year 2000.[11]

The alignment between residues 534-676 of SpCas9 and analogous amino acids of FnCas9 as published in Suppl. File 2 by Panda and Ray (2022) is reproduced below.

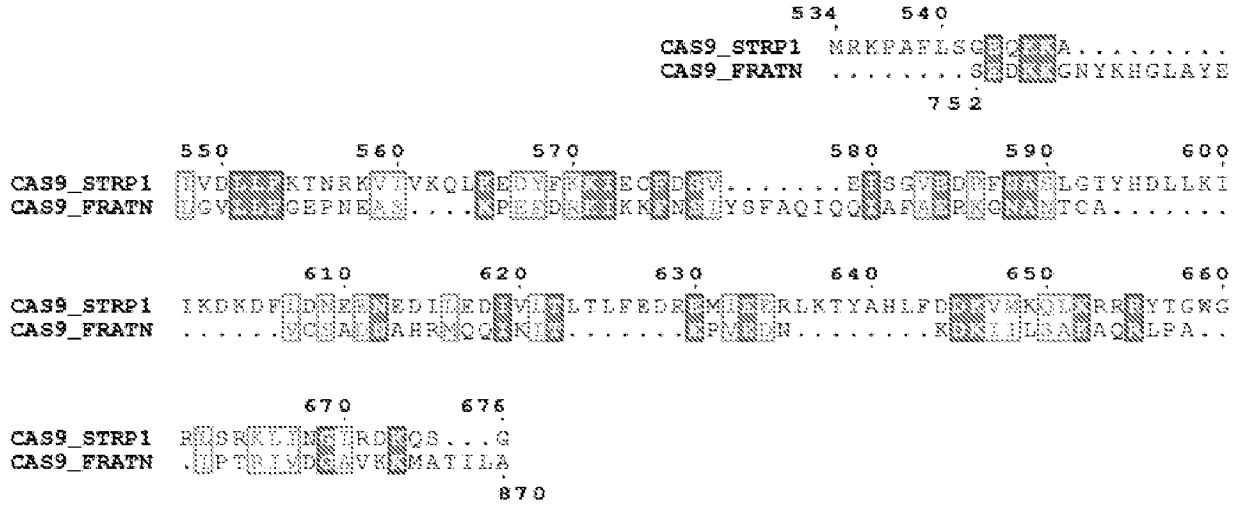
---

[8] amino acid sequence corresponding to residues 534–676 of SpCas9 are shown in bold (see sequence alignment)

[9] residues 534–676 are shown in bold.

[10] Panda G, Ray A. Comparative structural and dynamics study of free and gRNA-bound FnCas9 and SpCas9 proteins. Comput Struct Biotechnol J. 2022 Aug 4;20:4172-4184. doi: 10.1016/j.csbj.2022.07.041. PMID: 36016716; PMCID: PMC9389198.

[11] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000 Jun;16(6):276-7. doi: 10.1016/s0168-9525(00)02024-2. PMID: 10827456.



Applicant concludes that the recitation of "one or more sites on the FnCas9 that are analogous to the amino acid positions 534–676 of the *Streptococcus pyogenes* Cas9 (SpCas9) Rec1 domain" of claim 1, and claims dependent therefrom, would have been understood by a person of ordinary skill in the art in view of Nishimasu et al. (2014), the publicly available FnCas9 and SpCas9 amino acid sequences and the EMBOSS sequence alignment software.

Hence, claims 1, 3, 9–11, 13, 16, 35–39, 41, 49, 57, 59, 60, 62–67, 69-70, and 72 comport with the requirements of 35 U.S.C. § 112(b). Applicant respectfully requests withdrawal of the rejection.

Claim Rejections Under 35 U.S.C. § 112(a) – Written Description

Claims 1, 3, 9–11, 13, 16, 35–39, 41, 49, 57, 59-60, 62–67, 69-70 stand rejected, and new claim 72 is rejected under 35 U.S.C. § 112(a) as allegedly failing to comply with the written description requirement. The Office finds that the claims contain subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. *Office Action* at p. 10 (*citing Office Action* of July 3, 2023, at pp. 7–9).

Applicant traverses this rejection.

Claim 1

Claim 1, b) (ii) is amended to require:

(ii) one or more heterologous functional domains positioned at one or more sites ~~on the~~ in ~~FnCas9~~ FnCas that are analogous to [the] amino acid positions 94–

179 and 308–713 ~~534-676~~ corresponding to ~~of the~~ a *Streptococcus pyogenes* Cas9 (SpCas9) Rec1 domain, . .

Written description support for this amendment can be found, for example, in the following paragraphs of Provisional Patent Application 62/096,324, filed 23 December 2014:

> **[0058]** Due to crystal structure experiments, the Applicant has identified that ***positioning the functional domain in the Rec I domain***, the Rec2 domain, the HNH domain, or the PI domain of the SpCas9 protein ***or any ortholog corresponding to these domains*** is advantageous. . . .. (Emphasis added)
>
> **[00241]** In some embodiments, the CRISPR enzyme is part of ***a fusion protein comprising one or more heterologous protein domains*** (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). (Emphasis added)
>
> **[0260]** The invention comprehends optimized functional CRISPR-Cas enzyme systems. In particular the CRISPR enzyme comprises one or more mutations that converts it to a DNA binding protein to which ***functional domains*** exhibiting a function of interest ***may be*** recruited or appended or ***inserted*** or attached. (Emphasis added)

As noted above, Nishimasu (2014)[3] reported that the Rec1 domain of SpCas9 corresponds to amino acid **residues 94–179 and 308–713**.

Claim 72

Amended claim 72 requires:

> 72. (Currently amended) The vector system according to claim 1, wherein the FnCas protein or ortholog, homolog, or fragment thereof, comprises one or more heterologous functional domains positioned at one or more sites in FnCas that are analogous to amino acid positions 180-307 of a *Streptococcus pyogenes* Cas9 (SpCas9) ~~at one or more amino acid positions 175–306 of the FnCas~~ Rec2 domain.

Written description support for this amendment can be found, for example, in paragraph [0058] of the Provisional Patent Application 62/096,324, filed 23 December 2014 as shown below:

> **[0058]** Due to crystal structure experiments, the Applicant has identified that ***positioning the functional domain in*** the Rec I domain, ***the Rec2 domain***, the HNH domain, or the PI domain of the SpCas9 protein ***or any ortholog corresponding to these domains*** is advantageous. . . .. (Emphasis added)

As noted above, Nishimasu (2014)[3] reported that the Rec2 domain of SpCas9 corresponds to **amino acid residues 180-307**.

Claim 73

New claim 72 requires:

> 73. (New) The vector system according to claim 1, wherein the FnCas protein or ortholog, homolog, or fragment thereof, further comprises one or more heterologous functional domains positioned at one or more sites in FnCas that are analogous to amino acid positions 534-676 of a *Streptococcus pyogenes* Cas9 (SpCas9) Rec1 domain.

Written description support for new claim 73, which is dependent on claim 1, can be found, for example, in paragraph [0058] and [0251] of the Provisional Patent Application 62/096,324, filed 23 December 2014 as shown below:

> **[0058]** Due to crystal structure experiments, the Applicant has identified that ***positioning the functional domain in the Rec I domain***, the Rec2 domain, the HNH domain, or the PI domain of the SpCas9 protein ***or any ortholog corresponding to these domains*** is advantageous. . . .. (Emphasis added)
>
> **[0251]** Moreover, the crystal structure demonstrates that there is ***a flexible loop between approximately CRISPR-Cas9 (S. pyogenes) residues 534-676 which is suitable for attachment of a functional group*** such as an activator or repressor. (Emphasis added)

As shown in the sequence alignment (see p. 14 of this response), CRISPR-Cas9 (*Streptococcus pyogenes*) residues 534-676 are located within the Rec1 domain of SpCas9.

Furthermore, paragraphs [0058], [00251], [00255] and originally filed claims 31-32 of the Provisional Patent Application 62/096,324, filed December 23, 2014, refer to the crystal structure of SpCas9 and identify certain sites that may be suitable for **positioning of functional domains to the Rec1 domain** at position 553, Rec1 domain at 575, . . of the SpCas9 protein **or any ortholog corresponding to these domains** . . .." Thus, the '324 Provisional application describes the positioning of a heterologous functional domain within the region of amino acids 534–676 of SpCas9 referenced in pending claim 73. Paragraphs [00251] and [00255] teach the crystal structure data of SpCas9 may be used in combination with well-known homology modeling techniques as "a means to design modified CRISPR-Cas9s, such as by attaching thereto a functional group."

When a disclosure describes a claimed invention in a manner that permits one skilled in the art to reasonably conclude that the inventor possessed the claimed invention the written description requirement is satisfied. MPEP §2163. This possession may be shown in any number of ways and an Applicant *need not describe every claim feature exactly* because there is no *in haec verba* requirement. MPEP § 2163. Rather, to satisfy the written description requirement, all that is required is "reasonable clarity." MPEP § 2163.02[12] (Emphasis added). Also, an adequate description may be made in any way through express, implicit, or even inherent disclosures in the application, including words, structures, figures, diagrams, and/or formulae. MPEP §§ 2163(I), 2163.02. One of ordinary skill in the art would have reasonably concluded the inventor had possession of the claimed invention, as the disclosure satisfied the "reasonable clarity" requirement. Based on the totality of the disclosure, discussed in part *supra*, a person skilled in the art would be able to readily ascertain the sites on FnCas9 analogous to amino acids 534–676 of SpCas9 for positioning of the one or more heterologous functional domains as demonstrated by the sequence alignment between SpCas9 and FnCas9 described above.

Applicant submits that the claims comport with the requirements of 35 U.S.C. § 112(a) as the subject matter of claims 1, 72, and 73 is fully supported by the written description and a person skilled in the art, upon reading the specification, would be able to reasonably ascertain the corresponding FnCas amino acid positions at which to position the heterologous function domain. As such, Applicant respectfully requests reconsideration and withdrawal of the rejection of claims 1, 3, 9–11, 13, 16, 35–39, 41, 49, 57, 59, 60, 62–67, 69, and 70 under 35 U.S.C. § 112(a).

No waiver

All of Applicants' arguments and amendments are without prejudice or disclaimer. Applicants may not have addressed each specific rejection of the independent and dependent claims because Applicants submit that the independent claims are allowable, as discussed above. Applicants have not acquiesced to any such rejection and reserves the right to address the patentability of any additional claim features in the future.

[12] The fundamental factual inquiry is whether the specification conveys ***with reasonable clarity*** to those skilled in the art that, as of the filing date sought, the inventor was in possession of the invention as now claimed. See, e.g., *Vas-Cath, Inc.*, 935 F.2d at 1563-64, 19 USPQ2d at 1117. (MPEP § 2163)
(Emphasis added)

CONCLUSION

Applicant submits the foregoing as a full and complete response to the Office Action dated December 19, 2023, and that this Response places the application in condition for allowance and respectfully requests such action. If any issues exist that can be resolved with an Examiner's Amendment or a telephone conference, please contact the Applicant's representative at 912.257.4864.

The shortened statutory period expired on March 19, 2024. Accordingly, a petition for a one-month extension of time is submitted herewith that extends the time to file a Response from March 19, 2024 to April 19, 2024. No other fee is required for consideration of this paper. However, the Commissioner is hereby authorized to charge any fee deemed necessary for consideration of this paper, including any extension of time fees or additional claim fees, or to credit any overpayment, to Deposit Account Number 50-5193, reference number **BROD-3260US**.

JOHNSON, MARCOU, ISAACS & NIX, LLC
PO Box 691
Hoschton, GA 30548
Tel: 912.257.4864
198696
CUSTOMER NUMBER
docketing@jmin.com

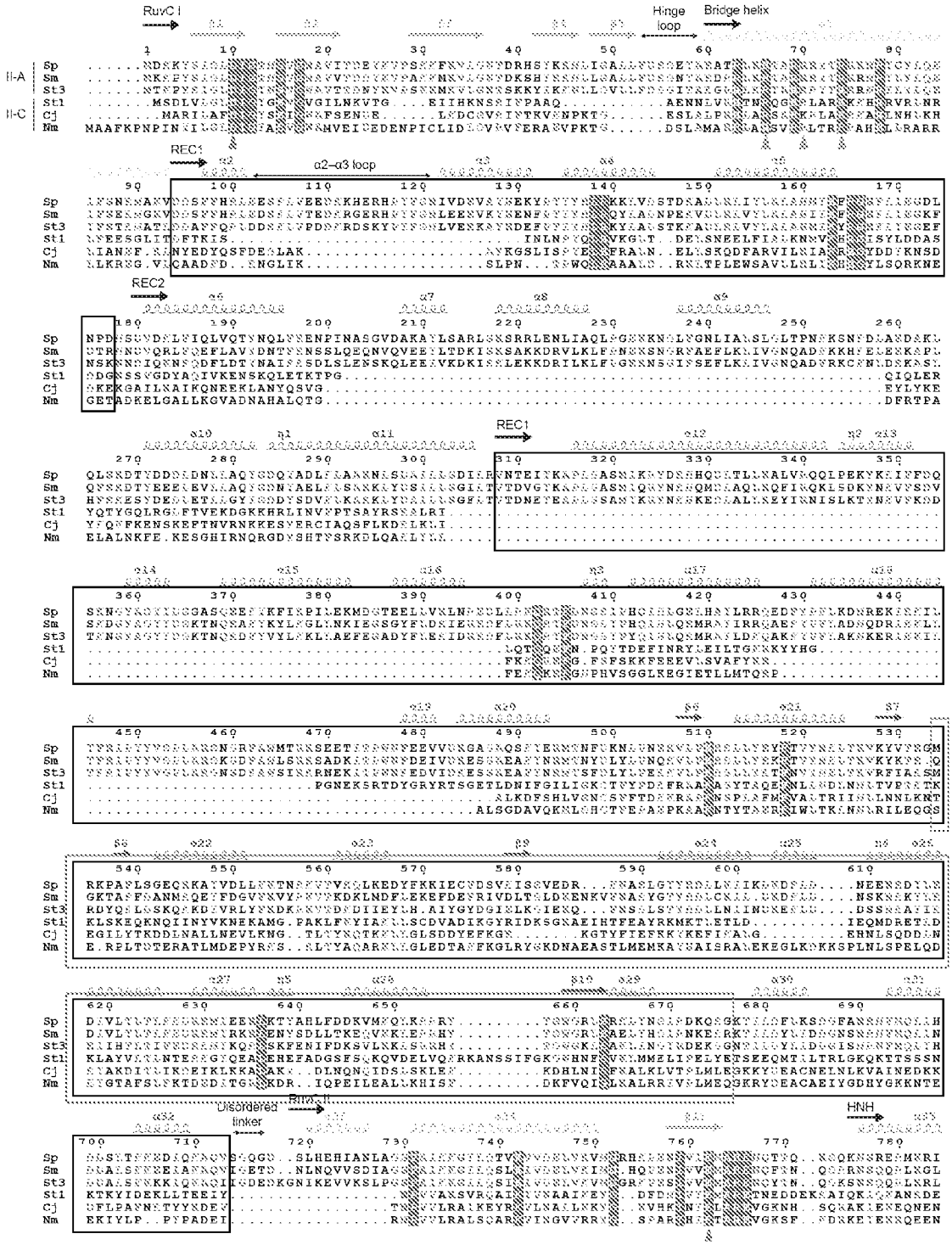
Respectfully submitted,

/Christopher Southgate/

Attorney for Applicant

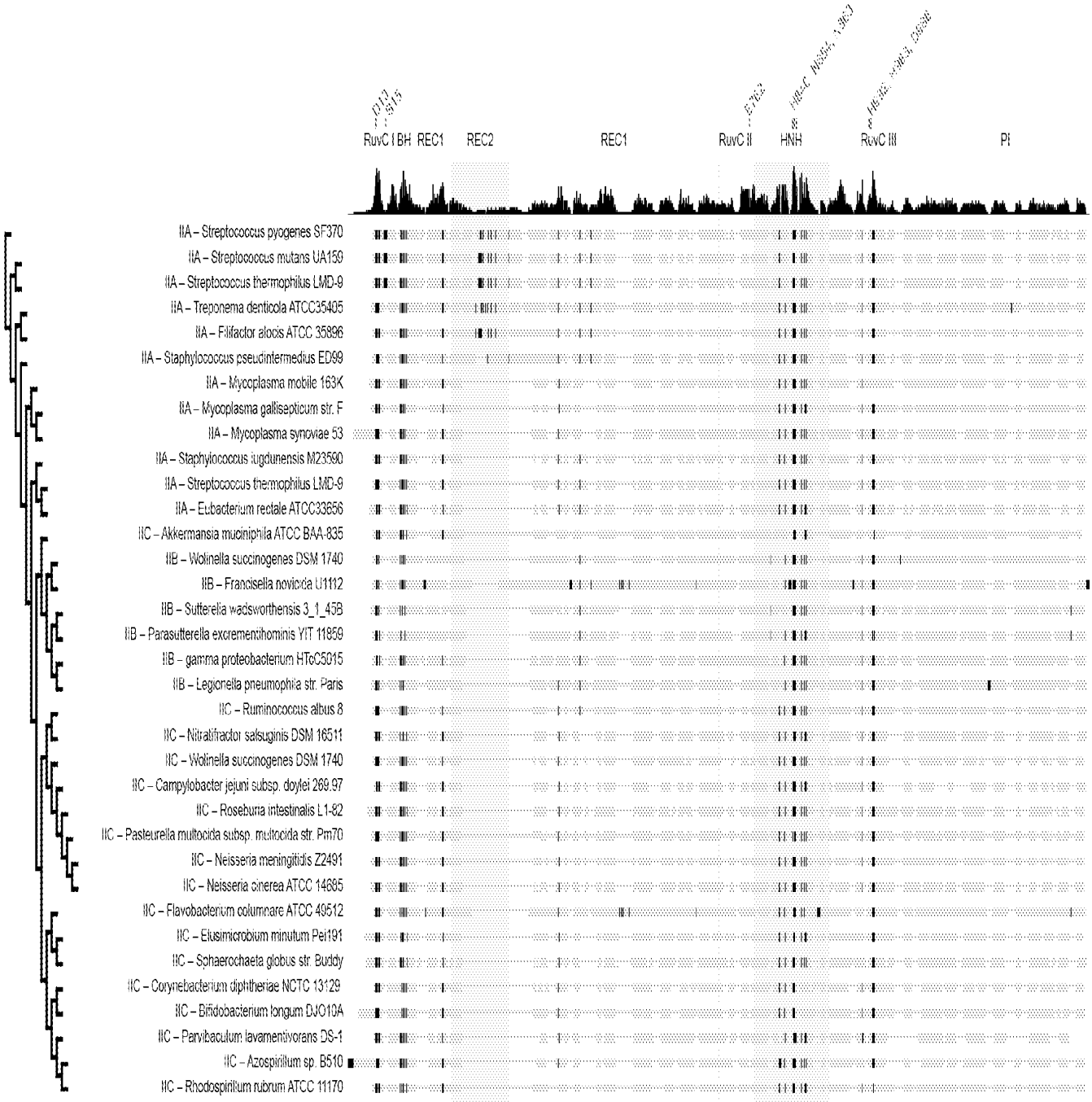
Christopher D. Southgate, Ph.D., Esq.
Reg. No. 54,875

APPENDIX IA

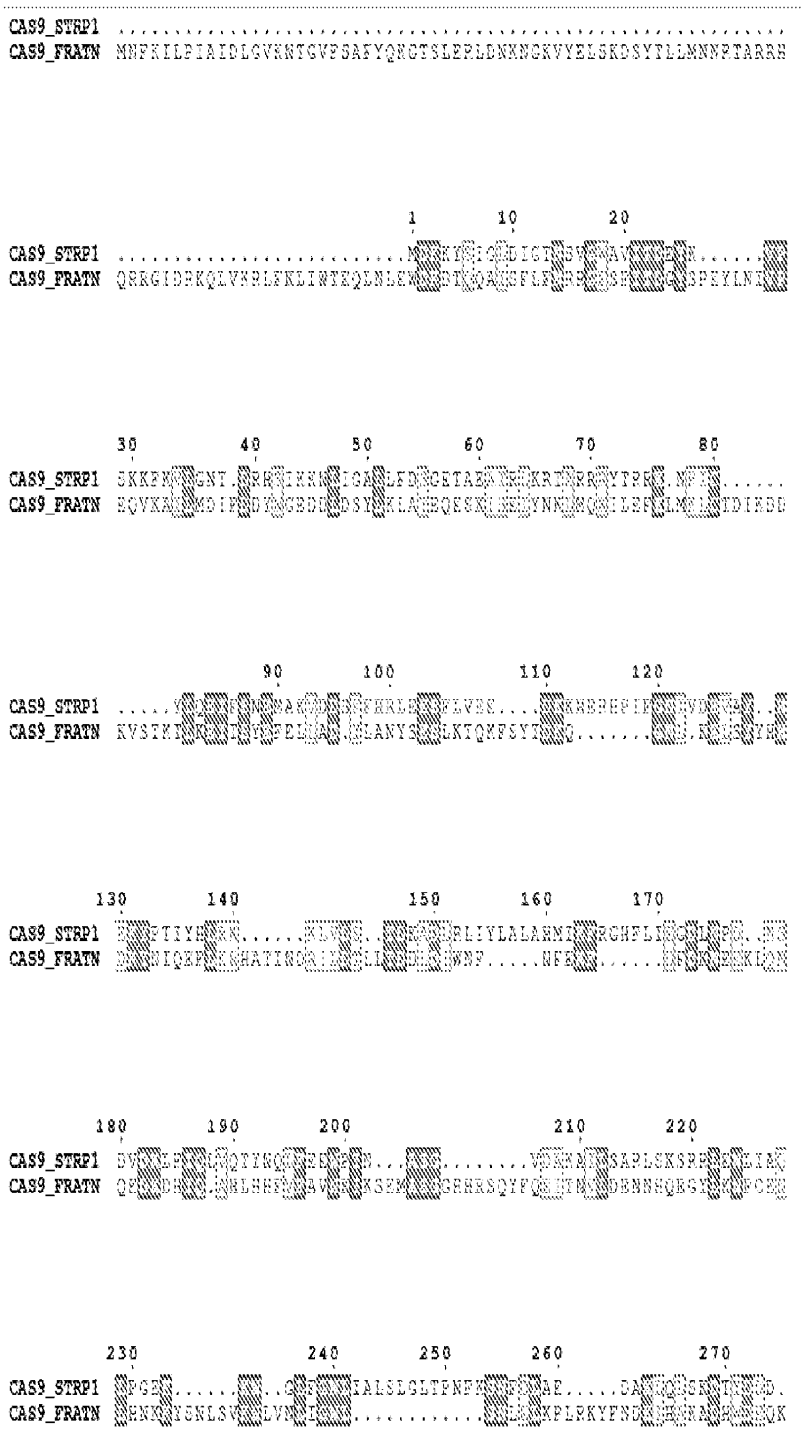


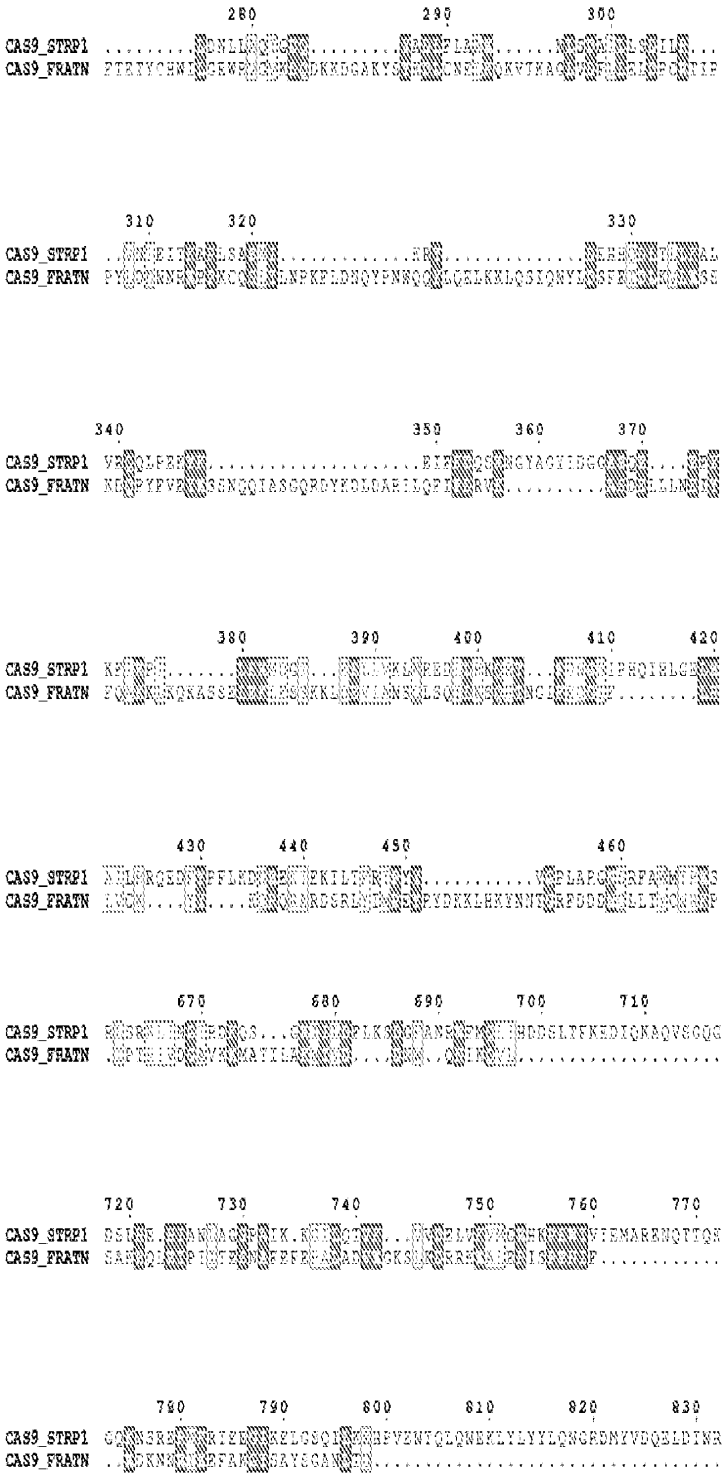
The catalytic residues are indicated by red triangles. Critical arginine residues on the bridge helix are indicated by green triangles. The secondary structure of S. pyogenes Cas9 is shown above the sequences. The figure was prepared using ClustalW (McWilliam et al., 2013) and ESPript (Gouet et al., 2003). Sp, S. pyogenes; Sm, Streptococcus mutans; St3, Streptococcus thermophilus CRISPR-3; St1, Streptococcus thermophilus CRISPR-1; Cj, Campylobacter jejuni; Nm, Neisseria meningitidis.

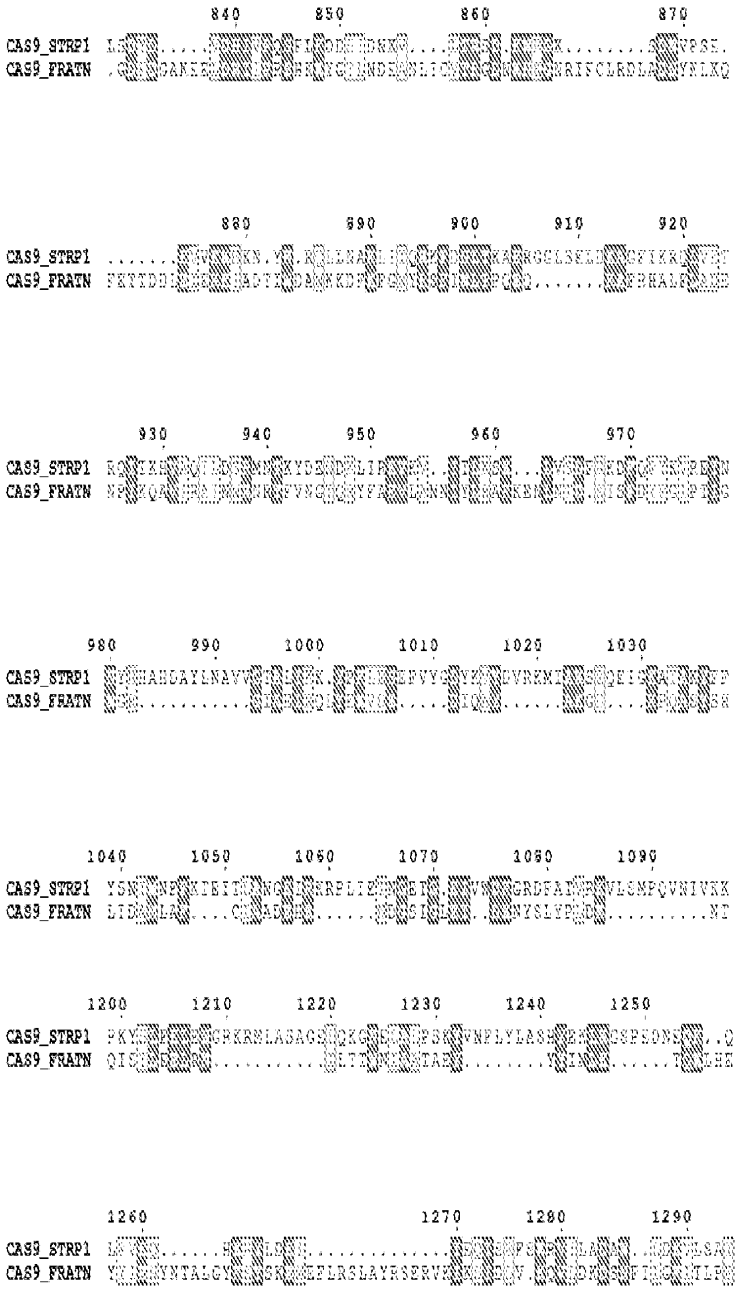
APPENDIX IB

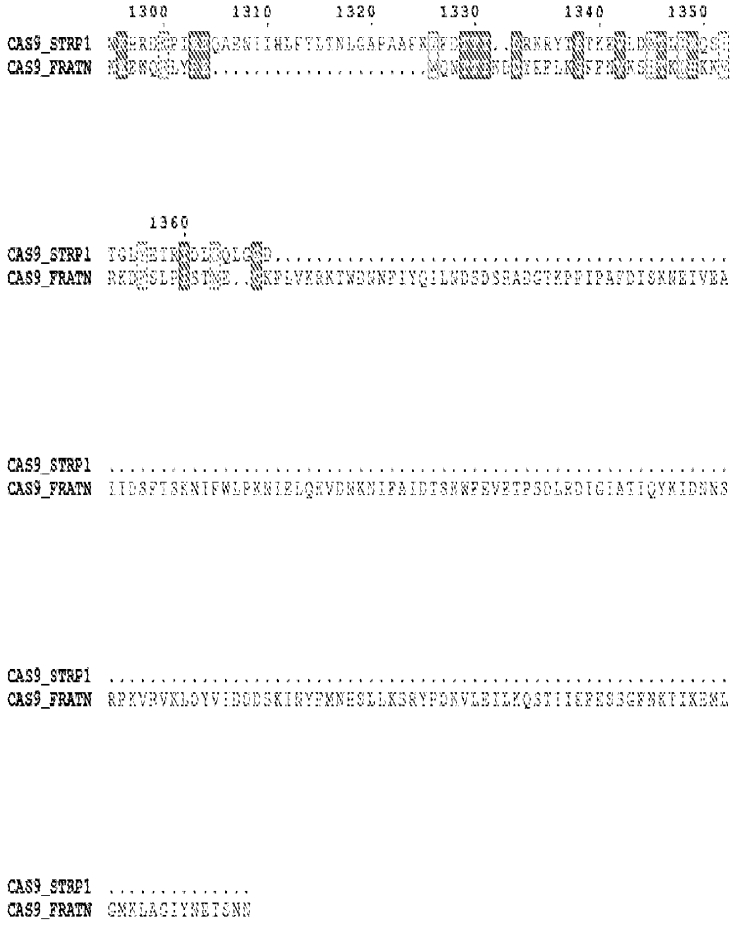

RuvC I
BH
REC1
REC2
REC1
RuvC II
HNH
RuvC III
PI
IIA – Streptococcus pyogenes SF370
IIA – Streptococcus mutans UA159
IIA – Streptococcus thermophilus LMD-9
IIA – Treponema denticola ATCC35405
IIA – Filifactor alocis ATCC 35896
IIA – Staphylococcus pseudintermedius ED99
IIA – Mycoplasma mobile 163K
IIA – Mycoplasma gallisepticum str. F
IIA – Mycoplasma synoviae 53
IIA – Staphylococcus lugdunensis M23590
IIA – Streptococcus thermophilus LMD-9
IIA – Eubacterium rectale ATCC33656
IIC – Akkermansia muciniphila ATCC BAA-835
IIB – Wolinella succinogenes DSM 1740
IIB – Francisella novicida U1112
IIB – Sutterella wadsworthensis 3_1_45B
IIB – Parasutterella excrementihominis YIT 11859
IIB – gamma proteobacterium HTCC5015
IIB – Legionella pneumophila str. Paris
IIC – Ruminococcus albus 8
IIC – Nitratifractor salsuginis DSM 16511
IIC – Wolinella succinogenes DSM 1740
IIC – Campylobacter jejuni subsp. doylei 269.97
IIC – Roseburia intestinalis L1-82
IIC – Pasteurella multocida subsp. multocida str. Pm70
IIC – Neisseria meningitidis Z2491
IIC – Neisseria cinerea ATCC 14685
IIC – Flavobacterium columnare ATCC 49512
IIC – Elusimicrobium minutum Pei191
IIC – Sphaerochaeta globus str. Buddy
IIC – Corynebacterium diphtheriae NCTC 13129
IIC – Bifidobacterium longum DJO10A
IIC – Parvibaculum lavamentivorans DS-1
IIC – Azospirillum sp. B510
IIC – Rhodospirillum rubrum ATCC 11170

APPENDIX II






CAS9_STRP1
CAS9_FRATN


1300
1310
1320
1330
1340
1350
1360
CAS9_STRP1
CAS9_FRATN

Doc code: RCEX
Doc description: Request for Continued Examination (RCE)

PTO/SB/30EFS (01-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

REQUEST FOR CONTINUED EXAMINATION(RCE)TRANSMITTAL
(Submitted Only via EFS-Web)

Application Number	15632067	Filing Date	2017-06-23	Docket Number (if applicable)	BROD-3260US	Art Unit	1633
First Named Inventor	Feng Zhang			Examiner Name	Maria Gomez Leavitt		

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.
Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, or to any design application. The Instruction Sheet for this form is located at WWW.USPTO.GOV

SUBMISSION REQUIRED UNDER 37 CFR 1.114

Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).

☐ Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.

- ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
- ☐ Other _____

☒ Enclosed

- ☒ Amendment/Reply
- ☒ Information Disclosure Statement (IDS)
- ☐ Affidavit(s)/ Declaration(s)
- ☐ Other _____

MISCELLANEOUS

☐ Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of months _____
(Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)

☐ Other _____

FEES

☒ **The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.**
The Director is hereby authorized to charge any underpayment of fees, or credit any overpayments, to Deposit Account No 505193

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

☒ Patent Practitioner Signature

☐ Applicant Signature

PTO/SB/30EFS (01-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Signature of Registered U.S. Patent Practitioner			
Signature	/Christopher Southgate/	Date (YYYY-MM-DD)	2024-05-17
Name	Christopher D. Southgate	Registration Number	54857

This collection of information is required by 37 CFR 1.114. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/SB/22 (12-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)

Docket Number (Optional): BROD-3260US

Application Number: 15/632,067

Filed: June 23, 2017

For: RNA-TARGETING SYSTEM

Art Unit: 1638

Examiner: Maria Gomez Leavitt

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
☐ One month (37 CFR 1.17(a)(1))	$220	$88	$44	$
☑ Two months (37 CFR 1.17(a)(2))	$640	$256	$128	$ 256
☐ Three months (37 CFR 1.17(a)(3))	$1,480	$592	$296	$
☐ Four months (37 CFR 1.17(a)(4))	$2,320	$928	$464	$
☐ Five months (37 CFR 1.17(a)(5))	$3,160	$1,264	$632	$

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

☐ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Director has already been authorized to charge fees in this application to a Deposit Account.

☐ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number _____.

☑ Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

☐ applicant/inventor.

☐ assignee of record of the entire interest. See 37 CFR 3.71. 37 CFR 3.73(b) statement is enclosed (Form PTO/SB/96).

☑ attorney or agent of record. Registration number 54,875.

☐ attorney or agent acting under 37 CFR 1.34. Registration number _____.

/Christopher Southgate/ — Signature

May 17, 2024 — Date

Christopher D. Southgate — Typed or printed name

912.257.4864 — Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

☐ * Total of _____ forms are submitted.

A Federal agency may not conduct or sponsor, and a person is not required to respond to, nor shall a person be subject to a penalty for failure to comply with an information collection subject to the requirements of the Paperwork Reduction Act of 1995, unless the information collection has a currently valid OMB Control Number. The OMB Control Number for this information collection is 0651-0031. Public burden for this form is estimated to average 6 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the information collection. Send comments regarding this burden estimate or any other aspect of this information collection, including suggestions for reducing this burden to the Chief Administrative Officer, United States Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450 or email InformationCollection@uspto.gov. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **If filing this completed form by mail, send to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. The United States Patent and Trademark Office (USPTO) collects the information in this record under authority of 35 U.S.C. 2. The USPTO's system of records is used to manage all applicant and owner information including name, citizenship, residence, post office address, and other information with respect to inventors and their legal representatives pertaining to the applicant's/owner's activities in connection with the invention for which a patent is sought or has been granted. The applicable Privacy Act System of Records Notice for the information collected in this form is COMMERCE/PAT-TM-7 Patent Application Files, available in the Federal Register at 78 FR 19243 (March 29, 2013). https://www.govinfo.gov/content/pkg/FR-2013-03-29/pdf/2013-07341.pdf

Routine uses of the information in this record may include disclosure to: 1) law enforcement, in the event that the system of records indicates a violation or potential violation of law; 2) a Federal, state, local, or international agency, in response to its request; 3) a contractor of the USPTO having need for the information in order to perform a contract; 4) the Department of Justice for determination of whether the Freedom of Information Act (FOIA) requires disclosure of the record; 5) a Member of Congress submitting a request involving an individual to whom the record pertains, when the individual has requested the Member's assistance with respect to the subject matter of the record; 6) a court, magistrate, or administrative tribunal, in the course of presenting evidence, including disclosures to opposing counsel in the course of settlement negotiations; 7) the Administrator, General Services Administration (GSA), or their designee, during an inspection of records conducted by GSA under authority of 44 U.S.C. 2904 and 2906, in accordance with the GSA regulations and any other relevant (i.e., GSA or Commerce) directive, where such disclosure shall not be used to make determinations about individuals; 8) another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)); 9) the Office of Personnel Management (OPM) for personnel research purposes; and 9) the Office of Management and Budget (OMB) for legislative coordination and clearance.

If you do not furnish the information requested on this form, the USPTO may not be able to process and/or examine your submission, which may result in termination of proceedings, abandonment of the application, and/or expiration of the patent.

Additional Uses

Additional USPTO uses of the information in this record may include disclosure to: 1) the International Bureau of the World Intellectual Property Organization, if the record is related to an international application filed under the Patent Cooperation Treaty; 2) the public i) after publication of the application pursuant to 35 U.S.C. 122(b), ii) after issuance of a patent pursuant to 35 U.S.C. 151, iii) if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections, or an issued patent, or iv) without publication of the application or patent under the specific circumstances provided for by 37 CFR 1.14(a)(1)(v)-(vii); and/or 3) the National Archives and Records Administration, for inspection of records.

PATENTS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No.	:	15/632,067	Conf. No.: 1196
First Named Inventor	:	Feng Zhang	
Filed	:	June 23, 2017	
TC/AU	:	1633	
Examiner	:	Maria Gomez Leavitt	
Customer No.	:	198696	
Docket No.	:	BROD-3260US	
Title	:	RNA-TARGETING SYSTEM	

INFORMATION DISCLOSURE STATEMENT

FILED VIA EFS-WEB

Commissioner:

Applicants cite the information on the attached Form PTO/SB/08, "Information Disclosure Statement by Applicant," pursuant to 37 C.F.R. §§ 1.56, 1.97, and 1.98.

No fee is due under 37 C.F.R. § 1.17(p) for this Information Disclosure Statement since it is being filed in compliance with 37 C.F.R. § 1.97(b)(4), before mailing of a first office action after the filing of an RCE under §1.114. However, the Commissioner is authorized to charge any fees deemed necessary for consideration of this paper to Deposit Account No. 50-5193 under attorney docket number BROD-3260US.

Certificate of EFS-WEB Transmission pursuant to 37 C.F.R. § 1.8

I hereby certify that this correspondence is being transmitted via the U.S. Patent and Trademark Office electronic filing system (EFS-Web) to the USPTO, on May 17, 2024.

/Christopher Southgate/

Christopher D. Southgate, Ph.D., Esq.

Reg. No. 54,875

Citation of the enclosed information does not constitute an admission of priority or that any cited item is available as a reference, or a waiver of any right the applicant may have under applicable statutes, Rules of Practice in patent cases, or otherwise.

JOHNSON, MARCOU, ISAACS & NIX, LL
PO Box 691
Hoschton, GA 30548
Tel: 912.257.4864
198696
CUSTOMER NUMBER
docketing@jmin.com

Respectfully submitted,

/Christopher Southgate/

_____
Attorney for Applicant

Christopher D. Southgate, Ph.D., Esq.
Reg. No. 54,875

PTO/SB/08a (03-15)
Approved for use through 07/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	15632067
Filing Date	2017-06-23
First Named Inventor	Feng Zhang
Art Unit	1638
Examiner Name	Maria Gomez Leavitt
Attorney Docket Number	BROD-3260US

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15632067
	Filing Date	2017-06-23
	First Named Inventor	Feng Zhang
	Art Unit	1638
	Examiner Name	Maria Gomez Leavitt
	Attorney Docket Number	BROD-3260US

	1	Panda G, Ray A. Comparative structural and dynamics study of free and gRNA-bound FnCas9 and SpCas9 proteins. Comput Struct Biotechnol J. 2022 Aug 4;20:4172-4184	☐
	2	Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000 Jun;16(6):276-7	☐

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	15632067
Filing Date	2017-06-23
First Named Inventor	Feng Zhang
Art Unit	1638
Examiner Name	Maria Gomez Leavitt
Attorney Docket Number	BROD-3260US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Christopher D. Southgate/	Date (YYYY-MM-DD)	2024-05-17
Name/Print	Christopher D. Southgate	Registration Number	54875

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

EMBOSS: The European Molecular Biology Open Software Suite

Peter Rice
pmr@sanger.ac.uk

Ian Longden
il@sanger.ac.uk

***Alan Bleasby**
ableasby@hgmp.mrc.ac.uk

The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK CB10 1SA.
*HGMP-RC, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK CB10 1SD.

In the area of sequence analysis, biologists find that existing software developed in the early days of sequencing often falls short of their needs. The rapid increase in large-scale sequencing and the challenges of the post-genomic era lead to a need for the rapid development of new applications, or the enhancement of existing software. This has been limited by the need for a general purpose-framework for the academic development of sequence analysis software. A survey by Pitr in 1998 (http://www.hgmp.mrc.ac.uk/CCP11/newsletter/vol2_1/questionnaire/) found that most bioinformatics developers were using C and that, although only 18% were using a free software library, a further 54% were likely to do so in future.

There is a further need to integrate existing packages and databases more effectively than present methods are able to. These problems have been debated at length by the members of EMBnet[1], with the conclusion at the 1996 meeting in Helsinki that there should be a concerted effort to provide an integrated sequence analysis software suite.

The EMBOSS suite (http://www.sanger.ac.uk/Software/EMBOSS/) is the result of this effort. With over 100 applications (Table 1) and the capability to be run from the traditional command line, through a web browser, or under more advanced graphical user interfaces, EMBOSS is suited to the needs of both expert users and practising scientists. The package runs on all Unix platforms, and it has also been built for Microsoft Windows NT (R. Bruskiewich, pers. commun.).

TABLE 1. Some of the most popular EMBOSS applications*

Program group	Selected applications
Alignment	Local (matcher, water) and global (stretcher, needle) alignment
Coding regions	Synonymous codon use (syco), codon statistics (chips)
Comparison	Large sequence word comparisons (dottup, polydot, wordmatch) and alignment (supermatcher)
Composition	General (compseq), frequent words (wordcount), graphical representation (chaos)
CpG islands	Report (cpgreport) and plot (cpgplot)
DNA features	Repeats (einverted, etandem), DNA melting (dan)
Editing	General editing utilities (cutseq, splitter), features (maskfeat), etc.
Indexing	Database indexing (dbiflat, dbigcg, dbiblast)
Motifs	Searching prosite (patmatdb, motifsearch), prints (pscan), transfac (ttfscan), general patterns (fuzznuc, fuzzpro)
Multiple alignment	Interface to clustalw (emma), display (prettyplot), editing (mse)
Protein features	Functional motifs (antigenic, sigcleave), structural motifs (pepcoil, helixturnhelix), amphipathic regions (pepnet, pepwheel), transmembrane prediction (tmap) and display (topo)
Protein properties	Hydropathy (pepwindow, pepwindowall, octanol), protease sites (digest), general (pepstats)
Sequence formats	Sequence reading/writing/format conversion (seqret, seqretall) and feature format conversion (seqretfeat)
Translation	Codon usage (cusp), reading frames (getorf, showorf, backtranseq)
Utilities	Motif database indexing (rebaseextract, ttextract, prosextract, printsextract), listing databases (showdb), searching for applications (wossname)

*For a full listing, see http://www.sanger.ac.uk/Software/EMBOSS/Apps/

TABLE 2. EMBOSS interfaces under development

Interface	Developer(s)	URL
W2H	EMBnet Germany	http://www.uk.embnet.org/embnet.news/vol4_1/w2gcg.html
www2gcg	EMBnet Belgium	http://www.uk.embnet.org/embnet.news/vol4_2/www2gcg.html
DisGUIse	EMBnet UK	http://www.hgmp.mrc.ac.uk/
PISE	Institute Pasteur	http://www.pasteur.fr/~letondal/Pise/ http://bioweb.pasteur.fr/seqanal/EMBOSS/
SRSWWW	EBI/Lion	http://srs.ebi.ac.uk/
ACEDB	Sanger Centre, UK and CNRS, Montpellier, France	http://www.acedb.org/
CINEMA	EMBnet Manchester, UK	http://www.umber.embnet.org/dbbrowser/
SeqPup	Indiana, USA	http://iubio.bio.indiana.edu/soft/molbio/seqpup/java
Staden	LMB, Cambridge, UK	http://www.mrc-lmb.cam.ac.uk/pubseq/
BioNavigator	EMBnet Australia and Encompass Pty	http://www.bionavigator.com/

Data sources

Sequence data can be used in any of the common sequence formats, with new formats easily added. For institutes where the public sequence databases are installed, EMBOSS can use these with a choice of formats, including GCG (Wisconsin Package Version 10, Genetics Computer Group, Madison, WI, USA; http://www.gcg.com/), SRS (Ref. 2), STADEN (http://www.mrc-lmb.cam.ac.uk/pubseq/), ACEDB (Ref. 3, see http://www.acedb.org/) and BLAST databases[4]. Failing this, individual sequence entries can be retrieved automatically over the web. Private local databases can also be used, for example by industrial users. Work is in progress on converting annotation (feature) formats.

The user interface

The basic interface is the command line. This is needed for many reasons, including the ability to run applications as part of larger automated analyses. The command line is defined in 'ACD' definitions (http://www.sanger.ac.uk/Software/EMBOSS/Acd/).

Most of the more friendly user interfaces in bioinformatics are simple forms which build the command line for the user, run the program, and present the results. EMBOSS is easy to convert to

0168-9525/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved. PII: S0168-9525(00)02024-2

FIGURE 1. EMBOSS

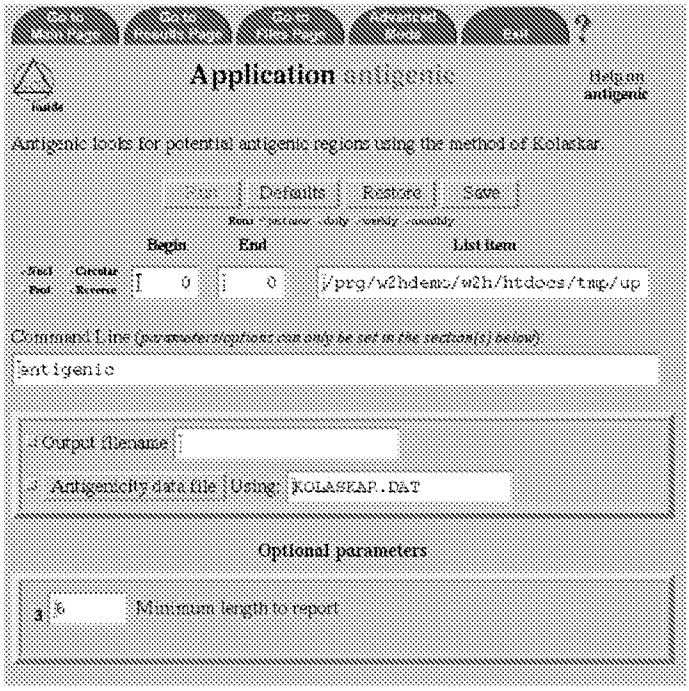
(a)

```
% antigenic
Finds antigenic sites in proteins
Input sequence(s): swissprot:act1_fugru
Minimum length [6]:
Output file: [act1_fugru.antigenic]: actin.anti
%
```

(b)

```
% antigenic swissprot:act1_fugru -out actin.anti -auto
```

(c)



trends in Genetics

(a) The command line interface, with prompts to the user. (b) Automated, with everything on the command line. (c) The W2H Web interface, generated automatically.

Applications

The original goal of EMBOSS was to re-implement applications in the 'EGCG package' (see http://www.uk.embnet.org/embnet.news/vol3_1/software.html) of extensions to GCG. This has now been achieved, with all the popular EGCG applications either replaced by EMBOSS or by other free software.

Additional applications have been contributed by several EMBnet members with their own development groups. The major contributors to date have been the UK node at Human Genome Mapping Project in Hinxton, Cambridge UK, with further programs from Norway, Germany, Italy, The Netherlands and the European Bioinformatics Institute. For most of the past year new applications have been added at the rate of ten per month. Most other EMBnet nodes are actively contributing to the project with interface developments, testing and documentation. The additional resources provided in this way are a typical benefit of open source software. Full documentation on all programs is available at http://www.sanger.ac.uk/Software/EMBOSS/Apps/.

Releases

The first developers' release was in the summer of 1998, together with an EMBnet workshop to review developments. The full beta-test release was made available in summer 1999, with a second EMBnet workshop where user feedback was gathered. We have implemented most of the recommendations from that workshop, and the first official release of EMBOSS followed in Spring 2000. The latest details are available from the EMBOSS Web pages (http://www.sanger.ac.uk/Software/EMBOSS/).

these interfaces by using the ACD definitions to generate the forms for each application automatically. Figure 1a, b shows an example application running from the command line, and Fig. 1c shows the web form for running the same program under the W2H web interface. Similar forms are available for other interfaces, including PISE (http://biowebpasteur.fr/seqanal/EMBOSS/). Table 2 lists the web and GUI interfaces we are currently collaborating with.

EMBOSS for developers

Fashions change in bioinformatics, and a number of programming languages have been favoured. In EMBOSS we have chosen to use standard C but we also support development in other languages. The first of these are from the Birkbeck Template Library (http://www.cryst.bbk.ac.uk/~classlib bioinf/BTL99.html). The licensing is open source (Gnu General Public Licence; http://www.gnu.org) so that EMBOSS is freely available. The software libraries are under the library GPL to allow other packages to link to EMBOSS. Examples so far include Phylip (http://evolution.genetics.washington.edu/phylip.html) and the MSE sequence editor.

References

1 Harper, R. (1997) EMBnet, an institute without walls. *Trends Biochem. Sci.* 21, 150–152
2 Etzold, T. and Argos, P. (1993) SRS, an indexing and retrieval tool for flat file data libraries. *Comp. Appl. Biosci.* 9, 49–57
3 Stein, D.L. (1999) Internet access to the *C. elegans* genome. *Trends Genet.* 15, 425–427
4 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
5 Rice, P. *et al.* (1996) EGCG 8.1 Release Notes. *Embnet.news* 3, 2–4

Computational and Structural Biotechnology Journal 20 (2022) 4172–4184

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Comparative structural and dynamics study of free and gRNA-bound FnCas9 and SpCas9 proteins

Gayatri Panda, Arjun Ray *

Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India

ARTICLE INFO

Article history:
Received 31 March 2022
Received in revised form 25 July 2022
Accepted 25 July 2022
Available online 4 August 2022

Keywords:
Conformational changes
Gene editing
off-target activity
Essential dynamics
Domain movement
Dynamic residues
Electrostatics

ABSTRACT

The introduction of CRISPR/Cas9 based gene editing has greatly accelerated therapeutic genome editing. However, the off-target DNA cleavage by CRISPR/Cas9 protein hampers its clinical translation, hindering its widespread use as a programmable genome editing tool. Although Cas9 variants with better mismatch discrimination have been developed, they have significantly lower rates of on-target DNA cleavage. Here, we have compared the dynamics of a more specific naturally occurring Cas9 from *Francisella novicida* (FnCas9) to the most widely used, SpCas9 protein. Long-scale atomistic MD simulation of free and gRNA bound forms of both the Cas9 proteins was performed, and their domain rearrangements and binding affinity with gRNA were compared to decipher the possible reason behind the enhanced specificity of FnCas9 protein. The greater binding affinity with gRNA, high domain electrostatics, and more volatility of FnCas9 than SpCas9 may explain its increased specificity and lower tolerance for mismatches.

© 2022 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

CRISPR/Cas (clustered regularly interspaced short palindromic repeat-CRISPR associated) systems confer viral resistance by recognizing and cleaving invading nucleic acids and have been found in the genomes of most archaea, and nearly half of bacteria [1,2]. The CRISPR/Cas immune response involves three stages – spacer acquisition/adaptation, crRNA expression-processing, and interference. The CRISPR/Cas system consists of a DNA cleaving enzyme, CRISPR-associated endonuclease (Cas protein), and its associated gRNA components (preceding the Cas operon), crRNA (CRISPR RNA), and tracrRNA (trans-activating gRNA, non-coding RNA links Cas9 with crRNA) which recognize the DNA cleaving region and initiate a double-stranded break.

The most often employed DNA cleaving enzyme from the type-II CRISPR/Cas system is SpCas9, which is derived from the bacteria *Streptococcus pyogenes*. SpCas9 enzyme is 1368aa long (4.1 kb) and adopts a bilobed architecture containing a nuclease lobe (NUC) and alpha-helical lobe or recognition (REC) lobe [3]. The nuclease lobe (residues 51–55 and 719–1368) comprise HNH (residues 780–906) and the RuvC-like domain residues (residues 1–55, 719–765, and 919–1099), Topoisomerase domain (residues 1099–1200) and C-terminal domain (residues 1201–1368). The HNH and RuvC domains help in DNA cleavage by bringing double-strand breaks into the DNA. The recognition lobe contains a 'bridge-helix' (residues 59–76) which connects the NUC and REC lobes with positively charged arginine residues [4,5]. Despite the rising use of the CRISPR/Cas9 technology in genome engineering, there are significant constraints that prevent it from being widely used. This includes Cas9's large size (which makes delivery difficult), recognition of a specific PAM (which limits its efficiency), and the introduction of random off-target mutations in target gene sequences. The CRISPR/Cas system's site-specific cleavage fails 15% of the time [6–8], resulting in unintended genetic alterations, and off-target effects. This limitation of Spcas9 led to the creation of novel Cas variants (e.g., Hifi Cas9, eSpCas9(1.1), SpCas9-HF1, SpCas9-NG, xCas9, Cpf1) with better efficiency [9–12]. The varying specificity and sensitivity of Cas variants emphasize the relevance of different Cas9 domains.

In work by Acharya et al., an ortholog of Cas9 from *Francisella novicida* (FnCas9), was shown to have very low non-specific editing compared to SpCas9 [13]. FnCas9 consists of 1,629 amino acids and is significantly larger than other Cas9 orthologs such as SpCas9 (1,368 amino acids) and SaCas9 (1,053 amino acids). FnCas9 comprises seven domains – the REC1-3, RuvC, HNH, WED (Wedge), and PI (PAM interacting) domains. The structural comparison of FnCas9 with SpCas9 revealed unanticipated conserved and divergent features of inter-domain interaction [14]. In contrast to *Streptococcus pyogenes* Cas9 (SpCas9), which has shown variable levels of off-

* Corresponding author.
E-mail address: arjun@iiitd.ac.in (A. Ray).

https://doi.org/10.1016/j.csbj.2022.07.041
2001-0370/© 2022 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

targeting due to tolerance of mismatches predominantly in the "non-seed" region of the gRNA, FnCas9 has been reported to have strong intrinsic specificity for its targets by tolerating only a single mismatch in the gRNA at the 5′ position of the PAM distal end [14]. Acharya et al. reported the specificity of FnCas9 is determined by its minimal binding affinity with DNA in the presence of mismatches, the staggered cleavage at the cleavage site, and higher homology-directed repair rates [13]. They have also suggested the basis behind the specificity of FnCas9 could be due to the interaction of its substrate with expanded REC3 and REC2 domains which are structurally different from those in SpCas9 and SaCas9 [14]. They also speculated that FnCas9's high electrostatic potential and interactions with several bases on PAM distal and proximal ends of the substrate are necessary for FnCas9 to achieve its cleavage competent state.

Extensive structural and mechanistic studies involving atomistic simulations, targeted-MD, and accelerated-MD provided interesting insights into the understanding of the recognition mechanism of SpCas9 domains, the mode-of-action, and the kinetics behind SpCas9 mediated DNA cleavage. In 2016, Palermo et al. performed all-atom molecular dynamic simulations of Apo-SpCas9, SpCas9-gRNA (binary complex) state, SpCas9-gRNA-tDNA bound to incomplete ntDNA and complete ntDNA state to reveal the conformational reorganization of Cas9 associated with nucleic-acid binding. They identified that the catalytic residue H840 was at a distance of 25 Å from the scissile phosphate of the tDNA in the absence of ntDNA, whereas in the presence of ntDNA, H840 stabilizes at a smaller distance of 15 Å from the scissile phosphate of the tDNA, suggesting that ntDNA binding is crucial for the formation of a catalytically competent Cas9 [15]. Gaussian Accelerated MD ($\sim$ 16 μs) on the pre-catalytic state of SpCas9 (PDB ID: 5F9R) showed how upon DNA-binding, the conformational dynamics and flexibility of HNH domain might facilitate the unwinding of dsDNA and formation of R-loop structure triggering the formation of active state [16,17]. Ab-initio MD study of the pre-catalytic state and active-state of SpCas9 was performed to disclose the two-metal-dependent mechanism of phosphodiester bond cleavage by RuvC Active site of CRISPR/Cas9 [18].

While much has been studied for the SpCas9 [15,17,18], the cleavage and recognition mechanism of FnCas9 and the factors responsible for its higher specificity are yet to be explored. Questions pertaining to the molecular inter-play for the interactions and a comparison between the two orthologs can provide us insights into the elusive future of designing customized Cas9 molecules.

To address these questions, crystal structures of gRNA-free and gRNA-bound forms of SpCas9 and FnCas9 complexes were simulated for 1 μs and analyzed. The principal component analysis, free energy landscape, and clustering methods were used to investigate the slow motions of Cas9 proteins and conformational changes due to the binding of gRNA. The relative domain movement was studied by distance plots between the center of mass of each domain pair. The crucial interactions and key residues involved in the binding of gRNA with Cas9 protein were examined at the Cas9 protein–RNA interfaces by binding free energy calculations.

2. Results and discussions

2.1. Conservation analysis reveals the diversity of Sp- and FnCas9 proteins.

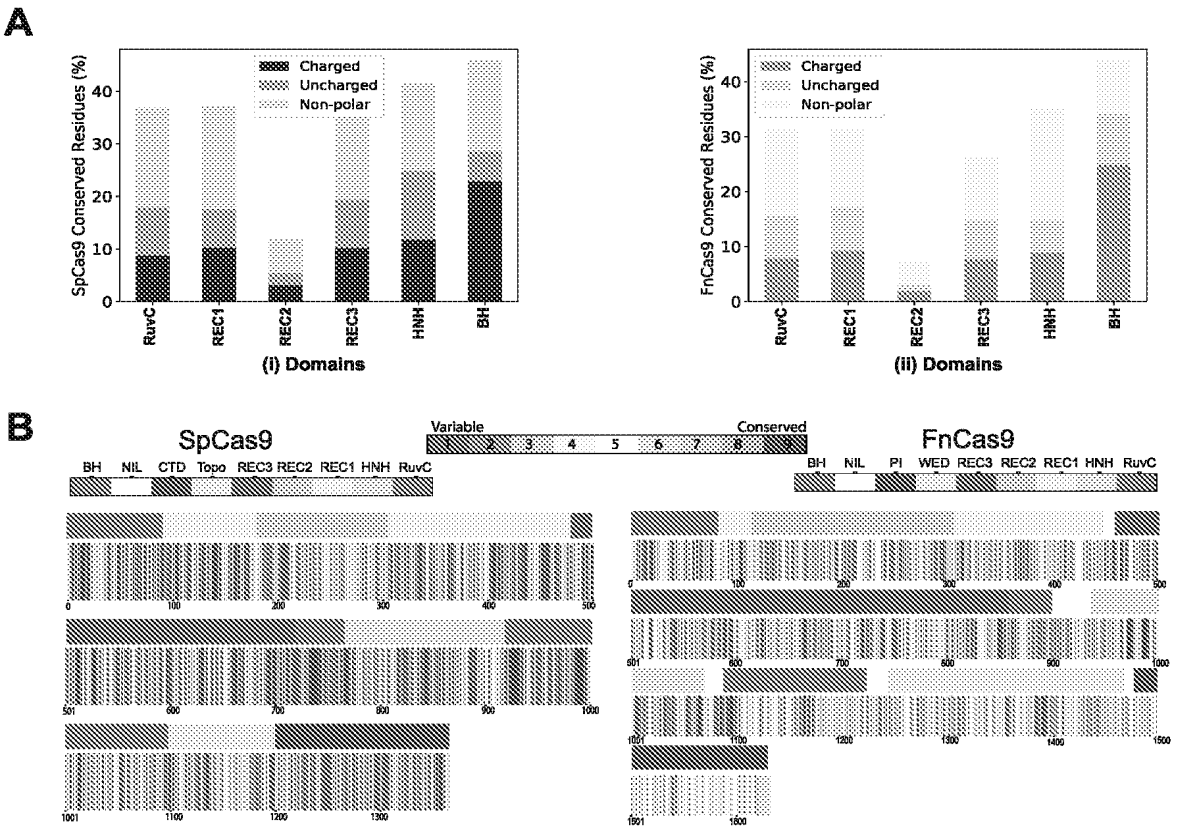
The results from ConSurf [19] revealed that bridge helix (BH) and the HNH domains were most conserved (SpCas9: 44% and 41%; FnCas9: 43% and 35%) and the REC2 domain was found to be least conserved (SpCas9: 12%; FnCas9: 7%) in both the proteins shown in Fig. 1A. In Fig. 1B, the conservation scores (Color codes) assigned by ConSurf were used to illustrate residue-wise and domain-wise conservation. Since the bridge-helix is an arginine-rich motif that bridges the nuclease and REC lobe, hence, the percentage of charged residues in this region is higher in both the proteins (SpCas9 (BH-Charged residues): 22.8% and FnCas9 (BH-Charged residues): 25%). In all the other domains, non-polar residues have shown a major involvement in conservation. Pairwise-sequence alignment of SpCas9 and FnCas9 sequences was performed using EMBOSS, which implements Needleman-Wunsch Algorithm to find the optimum alignment (including gaps) of two sequences along their entire length [20]. 23.9% residues in SpCas9 were conserved with the FnCas9 sequence (Supplementary file 2).

2.2. gRNA-binding to Cas9 proteins stabilizes the complex.

The atomic fluctuations and structural stability of the SpCas9 and FnCas9 systems during MD simulations were examined by monitoring the backbone root mean square standard deviation (RMSD) with respect to their starting structures as a function of simulation time. All the systems exhibited a stable deviation after 200 ns (200–1000 ns considered a stable part of the trajectory). Various geometrical properties, including the number of intra-protein hydrogen bonds, solvent accessible surface area (SASA), intermolecular hydrogen bonds between Cas9 protein and gRNA, and radius of gyration, were calculated from the equilibrium MD trajectories (Supplementary Figs. S1 and S2). Fig. 1C list the comparison between the average values of these geometrical properties in gRNA bound and unbound form for both the Cas9 systems. Time evolution of these properties was computed (Supplementary Figs. S1 and S2) revealing some subtle differences in geometrical properties in both cases. In the case of SpCas9, the average solvent accessibility (SASA) (Supplementary Fig. S1-C) and the average number of intra-protein hydrogen (Supplementary Fig. S1-D) bonds increased after gRNA-binding with a small increase in the compactness of the structure (Supplementary Fig. S1-B). The SASA plot (Supplementary Fig. S1-C) and average SASA values listed in (Table 1 in Fig. 1C) suggest an increase in solvent accessibility due to gRNA-binding leading to exposure of hydrophobic residues and loss of compactness/ folding of the protein. In the case of FnCas9, a similar observation was made with an increased radius of gyration (Supplementary Fig. S2-B) and SASA after gRNA binding (Supplementary Fig. S2-C), while the number of intra-protein hydrogen bonds (Supplementary Fig. S2-D) decreased after gRNA-binding. The average number of intermolecular hydrogen bonds between FnCas9 protein and gRNA (94-mer) (Supplementary Fig. S2-E) was more than in SpCas9 gRNA bound form (85-mer) Supplementary Fig. S1-E).

These observations suggest that gRNA-binding has stabilized the system (increased NHB); however, a reduced the compactness of the structure (increased ROG). This has led to increased solvent accessibility due to gRNA-binding and exposure to hydrophobic residues.

Residue-wise, RMSF is an often-used measure of conformational variance which highlights the regions of high mobility. To evaluate and compare the structural flexibility between free and gRNA bound forms of SpCas9 and FnCas9, per-residue root means square fluctuation (RMSF) was computed for the stable part of trajectory (200–1000 ns) and reported in Fig. 2A for SpCas9 and Fig. 3A for FnCas9. In Fig. 2A, we observe that the RMSF curve for Apo-SpCas9 (violet) and SpCas9 with gRNA (light-coral) had similar atomic fluctuations with average RMSF, 0.26 nm and 0.24 nm. These RMSF values were mapped onto the protein structure to find regions of higher flexibility. Regions in REC1, Topoisomerase, and RuvC-III domains were found to have higher fluctuation (RMSF > 0.6 nm) in unbound form (shown in Red color in Fig. 2B), whereas



C

Table 1: Geometrical Properties for stable part of SpCas9 trajectory (200-1000ns)

System	Average RMSD (nm)	Average RMSF (nm)	Average ROG (nm)	Average SASA(nm²)	Average-HBonds (count)	Intermolecular HBonds Spcas9-RNA
Apo-SpCas9	0.73	0.26	3.68	808.89	1053.22	90.613
SpCas9 + RNA	0.50	0.24	3.76	839.73	1108.82	

Table 2: Geometrical Properties for stable part of FnCas9 trajectory (200-1000ns)

System	Average RMSD (nm)	Average RMSF (nm)	Average ROG (nm)	Average SASA(nm²)	Average-HBonds (count	Intermolecular HBonds Spcas9-RNA
Apo-FnCas9	1.84	0.46	4.09	987.72	1224.58	131.12
FnCas9 + RNA	1.04	0.30	4.20	1007.67	1202.93	

Fig. 1. A. Grouped bar plot showing the percentage of conserved residues (residues with color codes ⩾8) in SpCas9 (PDB ID:4CMP) (on the left) and FnCas9 (PDB ID:5B2Q) (on the right) domains and the contribution made by charged, uncharged and non-polar residues in conservation. B. Residue and domain-wise conservation of residues in SpCas9 (on the left) and FnCas9 (on the right). The color schemes followed for domains and conservation color codes are added to the color bar on the top. C. Average geometrical properties calculated from stable part of MD trajectories for SpCas9 and FnCas9 systems.

only the Topoisomerase domain and some portions of CTD domains were found to be highly fluctuating in gRNA bound form of SpCas9, indicating lesser fluctuating regions post-gRNA-binding.

The atomic fluctuations in Apo-FnCas9 (indian-red) were slightly higher than the gRNA-bound form (teal) with average RMSF, 0.46 nm and 0.30 nm (Table 2 in Fig. 1C). The regions with high mobility (RMSF > 0.6 nm) were determined by mapping RMSF values onto the FnCas9 protein structure. It was found that regions in the REC lobe, specifically segment of REC2-REC3 domains (residues 218–233, 727–791), HNH (residues 950–998; 1004–1065) were highly fluctuating in the free-FnCas9 form (shown in red color) and only some regions of REC2 domain (residues 216–224), HNH domain (residues 953–961; 1004–1065), and PI domains (residues 1565–1572) had high-mobility in the gRNA bound form of FnCas9 as shown in Fig. 3. Both states of SpCas9 were equally flexible, whereas Apo-FnCas9 was relatively more flexible than its gRNA-bound form. On comparing, the radius of gyration, RMSD, and RMSF values, we can say that FnCas9 is more volatile than SpCas9 as it had a higher average radius of gyration and RMSF values.

Table 3
Computational alanine scanning result for hotspot residues involved in binding of gRNA with SpCas9 and FnCas9 protein. All values are in kcal/mol.

HotSpot Residues	$\Delta\Delta G_{(CAS)}$
Phe105 (SpCas9)	9.95 ± 1.2
Arg55 (FnCas9)	15.23 ± 1.5
Arg62 (FnCas9)	13.06 ± 0.8
Arg63 (FnCas9)	13.15 ± 1.8
Arg401 (FnCas9)	12.87 ± 1.7
Arg455 (FnCas9)	17.00 ± 1.7

2.3. Relative domain movement and domain structural deformations are involved in the binding of gRNA to Cas9 proteins

Many structural and biophysical studies have reported about the important conformational changes in Cas9 protein due to nucleic acid binding [21,15,17]. Upon DNA binding, the HNH domain undergoes a significant conformational change leading to activation of Cas9 protein for catalysis of TS [22,18]. Enhanced MD simulations of pre-activated CRISPR-Cas9 (5F9R.pdb) (~ 16 μs) revealed that the tight-coupling between REC lobe and HNH domain aided in bringing the Cas9 to the activated state. The conformational transitions observed in the REC lobe disclosed how the recognition domains 'sense' nucleic acids, 'regulate' the HNH conformational change, and ultimately 'lock' the HNH domain at the cleavage site, contributing to its catalytic activation [23]. These studies justify the significance of conformational dynamics of CRISPR-Cas9 in DNA catalysis, thereby improving its genome editing abilities.

The relative movements of intramolecular domains due to the binding of gRNA were determined by calculating distances between the center of mass of each domain as a function of simulation time for both the Cas9 forms.

The center of mass of HNH, RuvC, REC1, REC2, REC3, PI, and CTD domains were computed for both the systems (SpCas9 and FnCas9), and the distances between COM of every domain pair for the simulation length were calculated (Supplementary Fig. S3 (SpCas9) and S4 (FnCas9)). The domain movement due to gRNA binding was quantified by plotting a clustermap using the difference in average domain distance (200–1000 ns) in free and gRNA bound form. Upon gRNA binding in SpCas9, we observed that the domain movement by HNH-REC2 domains, CTD-Topo domains, and REC3-RuvC domains was similar (Fig. 4A). The HNH and REC2 domains moved closer to REC3- RuvC domains and moved

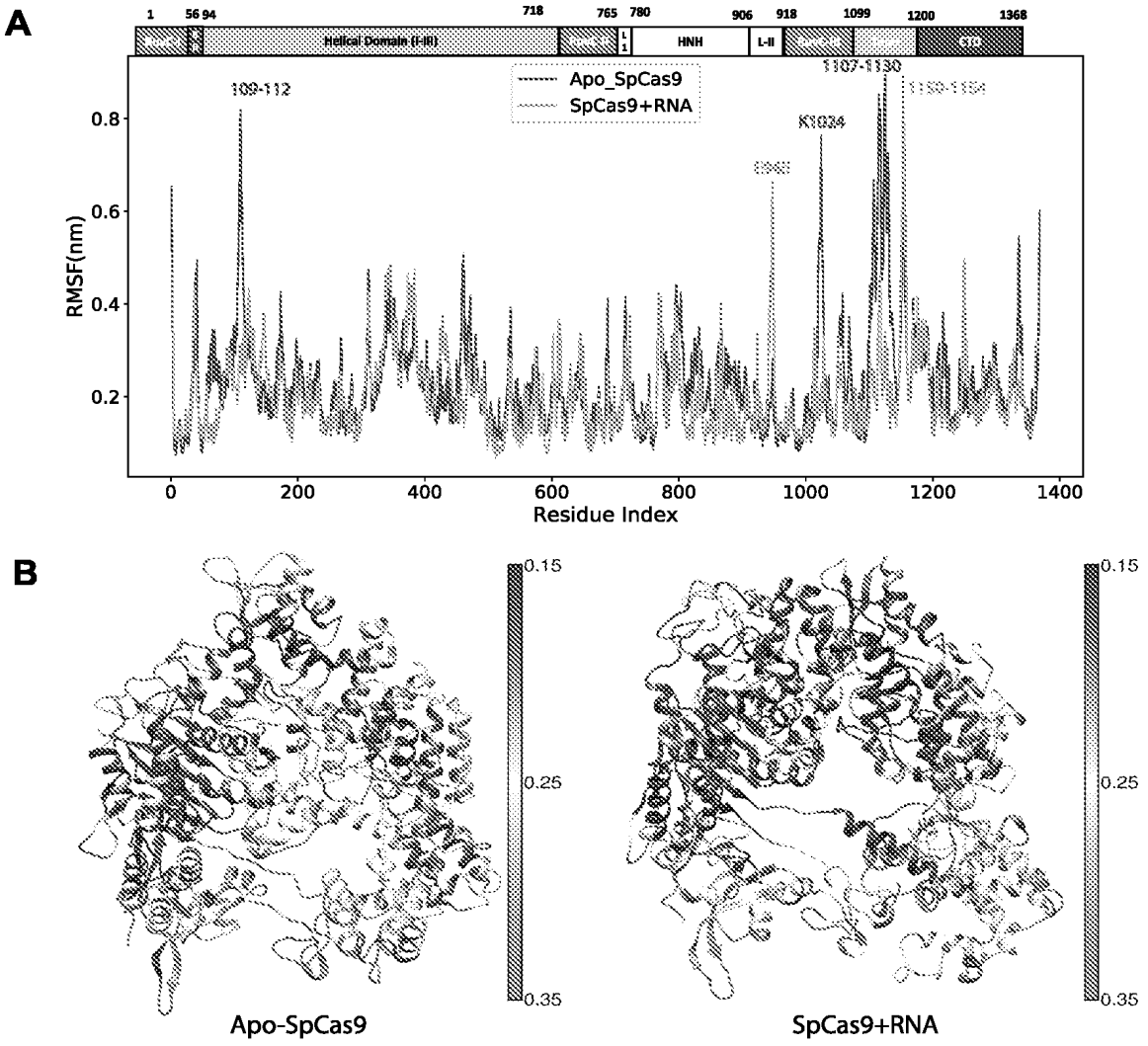


Fig. 2. Comparison between the structural flexibility of Apo-SpCas9 and SpCas9 with gRNA. (A) Per-residue RMSF profiles were calculated from MD trajectories of Apo-SpCas9 (violet) and SpCas9 + RNA (coral). The residues corresponding to every domain in the SpCas9 structure are shown. (B) The structural representations of Apo-SpCas9 and SpCas9 + RNA mapped with per-residue RMSF values generated using UCSF Chimera. The color ranges from blue to red and denotes that RMSF varies from the lowest to the highest values.

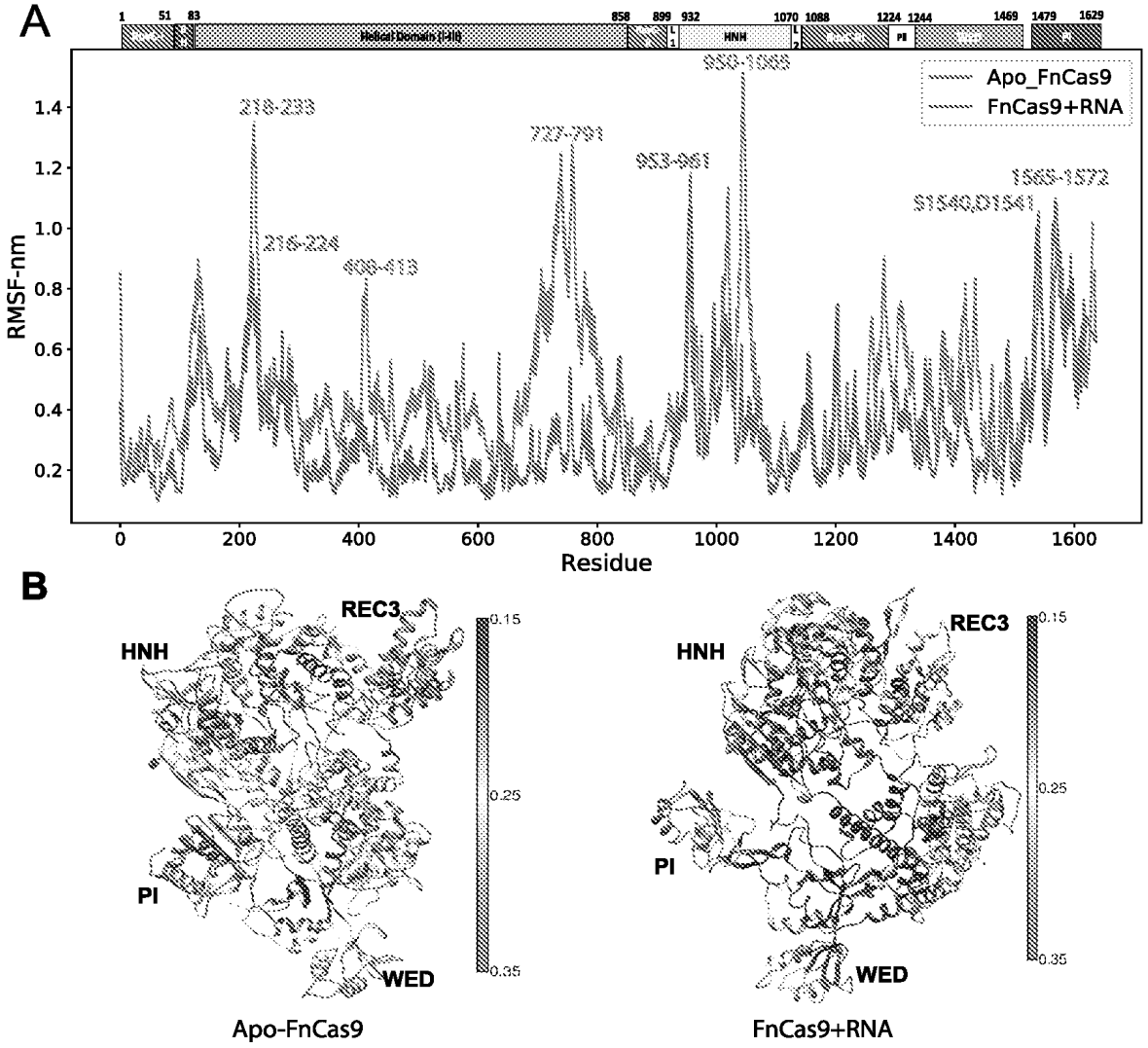


Fig. 3. Comparison between the structural flexibility of Apo-FnCas9 and FnCas9 with gRNA. (A) Per-residue RMSF profiles were calculated from MD trajectories of Apo-FnCas9 (Indian-red) and FnCas9 + RNA (teal). The residues corresponding to every domain in the FnCas9 structure are shown in the top horizontal bar. (B) The structural representations of Apo-FnCas9 and FnCas9 + RNA mapped with per-residue RMSF values generated using UCSF Chimera. The color ranges from blue to red and denotes that RMSF varies from the lowest to the highest values.

away from the REC1 domain after gRNA binding in SpCas9 form. Similarly, CTD and Topo moved closer to REC2, and REC3, and moved away from RuvC and REC1 domains. The RuvC and REC3 domains moved closer to REC2 and HNH domains and moved farther from Topo domains. In Fncas9, the PI, WED, and RuvC domains moved closer to each other, and REC3 and HNH domains, and moved away from the REC2 domain (Fig. 4B). REC1 and REC2 domains have shown a similar domain movement by moving away from RuvC, PI, and WED domains. The FnCas9 protein has shown larger domain deviations as compared to SpCas9, as the domain deviation scale was larger in FnCas9.

The structural deformation of each of these domains in terms of domain flexibility and domain-wise structural transitions was studied by computing the domain-wise RMSF (Fig. 4C and D). The atomic fluctuations of every domain in the free form of SpCas9 were observed to be greater than in the gRNA-bound form. This observation was in agreement with other observations (in Fig. 3 and Fig. 2), suggesting increased stability in the presence of nucleic acids. The Topo and REC1 domains were found to be highly mobile in both the forms of SpCas9. The domain-wise average RMSF was calculated for Apo-FnCas9 and FnCas9 with gRNA for the stable part of the trajectory and plotted in Fig. 4B. A similar observation was made in the case of FnCas9 too, where the atomic fluctuations of domains in the free form of FnCas9 were greater than the gRNA bound form for RuvC, REC1, and REC2 domains, whereas the average RMSF of the HNH domain was higher in gRNA bound form than in the free form. The PI and REC3 domains were highly mobile in the pre-gRNA-bound form of FnCas9.

To evaluate the conformational flexibility of these domains, the percentage of residues existing in coil conformation for more than 70% of the simulation time was calculated for every domain in Apo and gRNA-bound form for both the systems in SpCas9 form (Fig. 4C) and in FnCas9 form (Fig. 4D). In SpCas9, the percentage of residues existing as a coil for more than 70% times, decreased after gRNA binding for all the domains except in the HNH and the REC2 domains. In FnCas9, the percentage of coils increased after gRNA-binding in RuvC, REC1, and PI domains.

The propensity of a residue to exist as a helix, sheet, turn, or coil was computed for the stable part of the trajectory. Also, dynamic residues with high order of transition between secondary structures were annotated (See Materials and Methods for details). The secondary structure assignment of residues in Apo-SpCas9

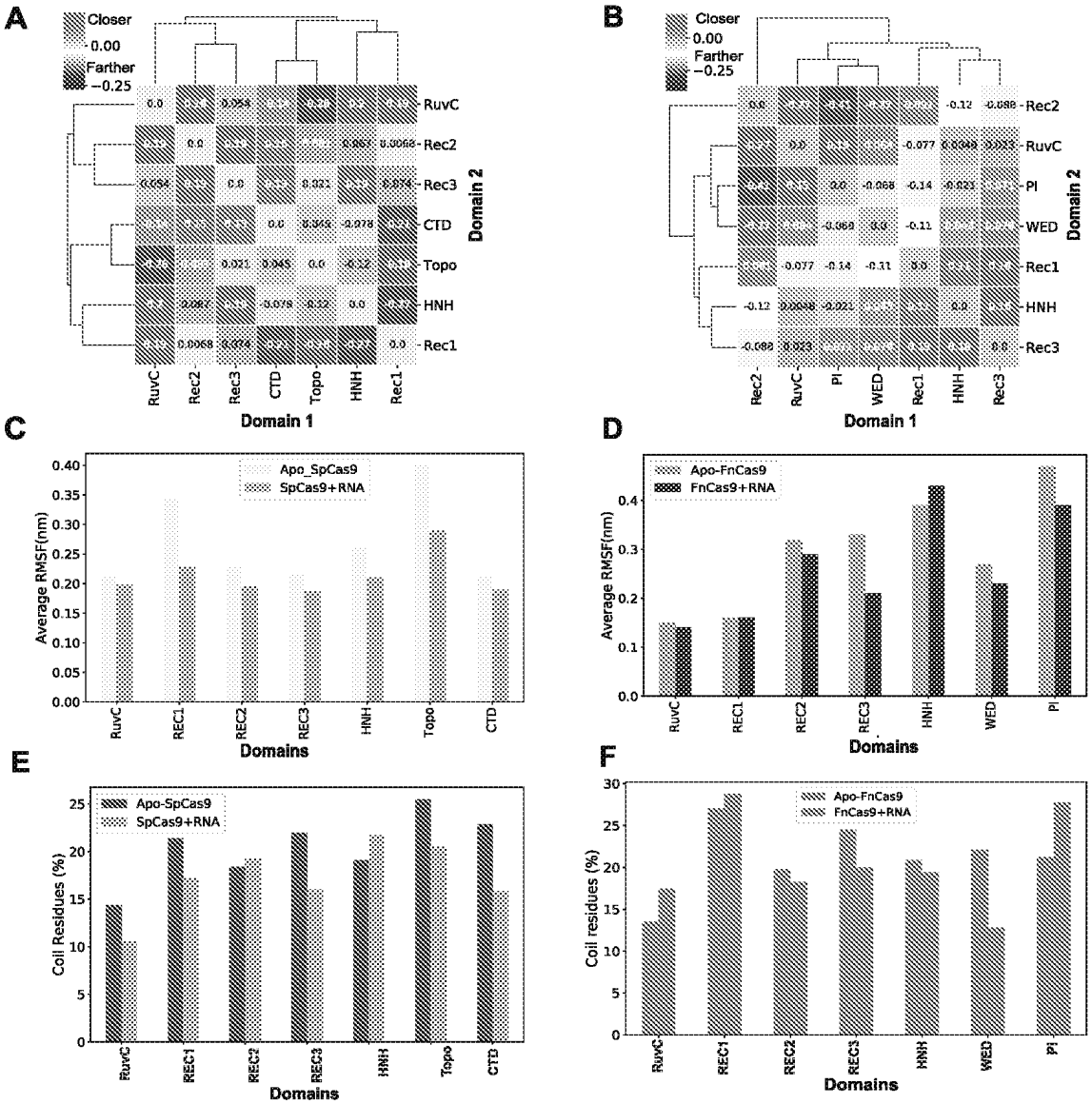


Fig. 4. Domain structural movements and transitions in clustermap shows domain movement between domains on the X and Y-axis. Red boxes suggest domains moved closer and blue boxes are for domains that moved away. The color intensity depends on the magnitude of domain movement. (A) Clustermap for SpCas9 (B) Clustermap for FnCas9 grouped bar plot with average RMSF values (200–1000 ns) corresponding to RuvC, REC1, REC2, REC3, and HNH domains in the free and bound form of (C) SpCas9 system and (D) FnCas9 system. Domain-wise percentage of Coil residues (>70% part of trajectory) in (D) SpCas9 system and (E) FnCas9 system.

and SpCas9 with gRNA are shown in Supplementary Fig. S5-A(i) and Supplementary Fig. S5-B(i), and Apo-FnCas9 and FnCas9 bound to gRNA are shown in Supplementary Fig. S7-A(i) and S7-B(i), respectively. The probability of residue to exist as a helix, sheet, or coil for 1000 frames was calculated and plotted to see regions of structural deformations at the residue level (SpCas9-Supplementary Fig. S5-A(ii) and S5-B(ii); FnCas9-Supplementary Fig. S7-A(ii) and S7-B(ii)). The secondary structure transition of residues among several conformations of proteins contributes to the structural flexibility of the proteins. The percentage of these dynamic residues that make up each of the functional domains of both the forms of SpCas9 and FnCas9 proteins are shown in Fig. 5A and Fig. 5B. The percentage of dynamic residues has increased after gRNA binding for all the domains in bot Cas9s. In the case of Spcas9, residues with higher structural transitions were found to predominantly exist in the REC lobe while the HNH domain had the least dynamic residues. The REC2 domain had shown a drastic increase in dynamic residues from 15% to 35% after gRNA binding, and on the contrary, HNH and CTD domains had shown less percentage of dynamic residues. This observation was in line with conservation analysis where the HNH domain was the most conserved, and the REC2 domain was least conserved. The conformational flexibility of a protein region also depends on the proportion of regions with NMR chemical shifts of a random-coil; regions that lack significantly ordered secondary structure (as determined by CD or FTIR) [24]. A similar observation was made in the case of FnCas9, the percentage of dynamic residues increased after gRNA binding. Here, the REC lobe contained the highest percentage of dynamic residues after gRNA binding.

After applying 70% cut-off, it was observed that the protein structure was predominantly constituted by helices (30 – 40% residues), followed by the dynamic residues (16 – 23% residues) and coils (16 – 20%), while sheets, turns, and bends were the least favored (4 – 12% respectively) (Supplementary Fig. S6). Interestingly, RMSD and geometrical properties of the system during the simulation revealed that the gRNA binding was stabilizing the

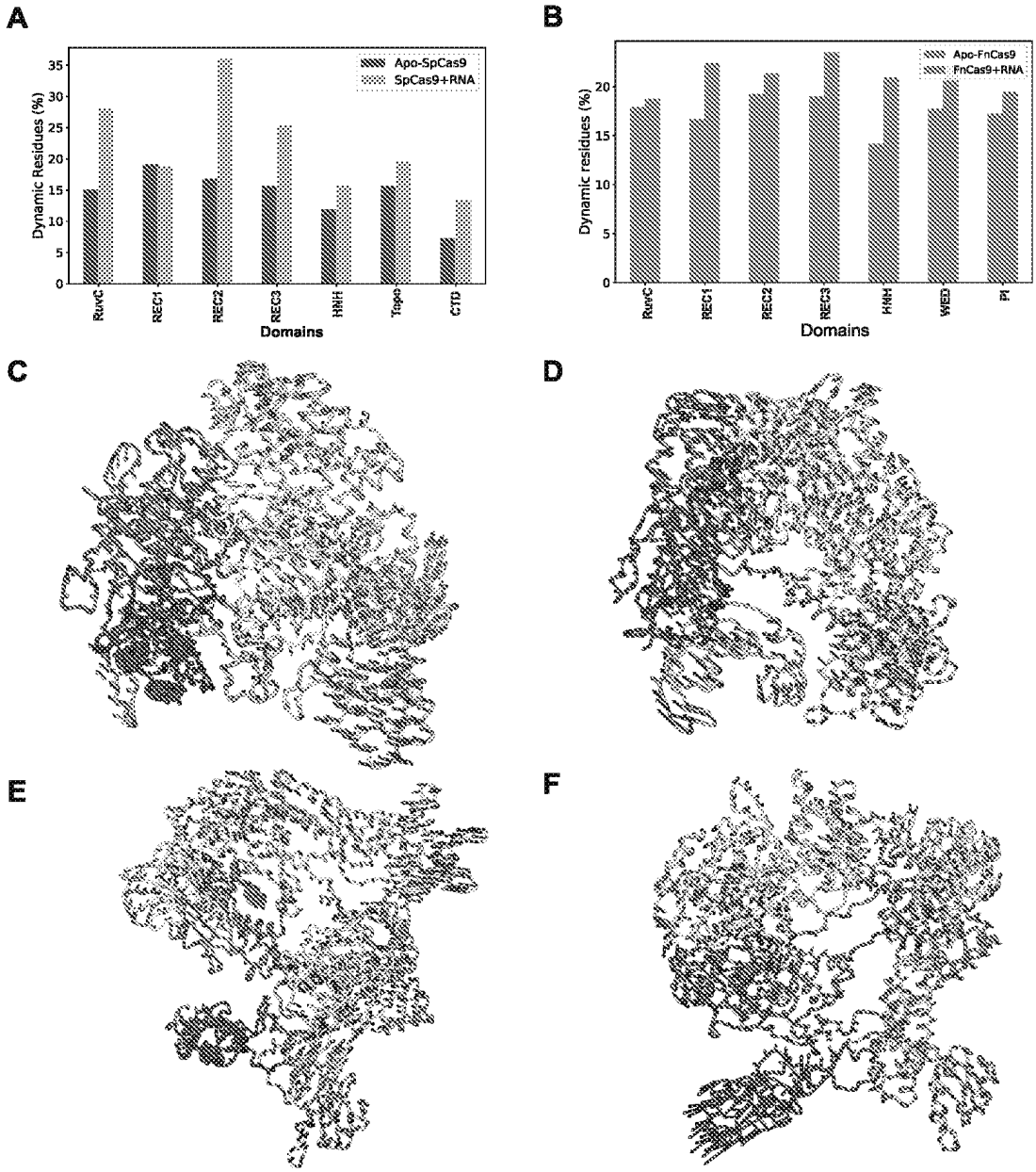


Fig. 5. Domain structural movements and transitions in FnCas9. Domain-wise percentage of dynamic residues (>70% part of trajectory) in (A) Apo-SpCas9 and SpCas9 with gRNA and in (B) Apo-FnCas9 and FnCas9 with gRNA. Most dominant mode, containing 100 conformations along eigenvector 1 for (C) Apo-SpCas9 and (D) SpCas9 with gRNA (E) Apo-FnCas9 (F) FnCas9 with gRNA form and visualized in NMWiz of VMD. The arrows indicate the direction and relative amplitude of motions.

structure, however, the secondary structure analysis revealed a decrease in percentage helicity (Apo-SpCas9: 42.03%, SpCas9 + RNA: 37.21%; Apo-FnCas9: 38.73%, FnCas9 + RNA: 30.95%) and increase in the percentage of dynamic residues (Apo-SpCas9: 16.74%, SpCas9 + RNA: 22.15%; Apo-FnCas9: 17.03%, FnCas9 + RNA: 21.39%) due to the binding of gRNA shown in Supplementary Fig. S6. The Topo and REC1 domains had higher residues as coils which could be a reason behind the high average RMSF/fluctuations (Supplementary Fig. S6).

2.4. Larger domain deviations are observed in SpCas9 protein than in FnCas9

The principal component analysis was performed to capture the essential motions of the simulated systems. To investigate the dominant motions of free and gRNA-bound forms of Cas9 along the first principle component (PC1), corresponding to the system's largest amplitude motion, 100 conformations of the system along eigenvector1 were stored in PDB format and visualized in Normal Mode Wizard of VMD [25]. The eigenvalues indicated fluctuations of the eigenvector in the hyperspace. Supplementary Fig. S8-B (SpCas9) and S8-D (FnCas9) show that only a few eigenvectors have larger eigenvalues which played a major role in the overall motion of the system.

The overlap between principal components and the coordinates of the trajectory was determined for free and bound states of SpCas9 and FnCas9. Supplementary Fig. S8-A and S8-C show the conformational sampling of free and bound forms of SpCas9 and FnCas9, respectively, in the essential subspace by projecting the backbone atom, showing the tertiary conformations along eigenvector 1 and 2. The 2D projection clearly depicted that free forms of Cas9 cover a wide range of phase spaces as compared to

gRNA-bound forms. The decrease in the overall motion of Cas9 forms due to gRNA binding indicates the stability of the structure in the presence of nucleic acids.

In Fig. 5C, the dynamics of Apo-SpCas9 protein along the first principal mode of motion (PC1) are shown. REC1, HNH, and CTD domains have shown larger domain movements. The RuvC domain moved closer to the REC lobe. The HNH and CTD domains have moved towards the REC1 domain, and moved away from REC3 and REC2 domains. The dominant motions in the case of the gRNA-bound form of SpCas9 in Fig. 5D, included REC1, CTD, and Topo domain movements. HNH and RuvC domains moved closer to the REC1 domain. The domain movements in the case of the free form of SpCas9 were larger than the gRNA-bound form.

The most dominant mode of motion of the Apo-FnCas9 form is shown in Fig. 5E. The HNH, RuvC, WED and REC3 domains moved closer to the REC2 domain. The WED domain moved away from HNH and PI domains. In gRNA bound form (Fig. 5F), major domain movement is shown by PAM interacting domain (PI domain), it moved away from the REC lobe. The reorganization of the PI domain after gRNA binding is indicative of the structural preparation of FnCas9 for the next step of gene editing, which requires scanning of PAM motif and binding with the DNA strands. In both the Cas9 forms, the large amplitude motions were observed in the Cas9 domains which are directly involved in the process of recognition and binding of nucleic acids, i.e. REC lobe, regions of HNH, PI, and CTD domains. The ribonucleic acid-binding has stabilized the complex. In the first mode of motion, SpCas9 protein has shown larger domain deviations as compared to FnCas9.

A three-dimensional free energy landscape (FEL) was built for the free and gRNA-bound form of both Cas9 systems using the projections of first (PC1) and second (PC2) eigenvectors to show protein stability in terms of Gibbs Free energy. Fig. 6 shows the FEL values for Apo and gRNA bound form of SpCas9 and FnCas9. In the case of Apo-SpCas9, there was only one main global free energy minimum region, indicating only one stable conformational state. The lowest energy conformer of Apo-SpCas9 was identified at 636 ns with local free-energy 0 kJ/mol (State A). The second low energy state in the same energy basin had the energy of 0.74 kJ/mol (State B). The inter-conversion from State A to State B had an energy barrier of 2.2 kJ/mol while from State B to State A had an energy barrier of 2.96 kJ/mol (Supplementary Fig. S9-A). Whereas the gRNA-bound form of SpCas9 had shown two low-energy basins with energy minima at 508 ns and 850 ns (G = 0 k J/mol). The inter-conversion between these states had an energy barrier of 3.8 kJ/mol (Supplementary Fig. S9-B). The state transition barrier is slightly higher in the gRNA-bound form than in the free-form of SpCas9, suggesting that state transition is easier in Apo form than in gRNA bound form of SpCas9. Likewise, in Apo-FnCas9 and FnCas9 bound to gRNA, two energy basins were observed. The lowest energy conformation was extracted at 478 ns and 249 ns, respectively. In Apo-FnCas9, the lowest energy state was at 0 kJ/mol (State A), and the second low energy state had 1.6 kJ/mol (State B) of energy. The transition from State A to State B of Apo-FnCas9 had an energy barrier of 5.9 kJ/mol and from State B to State A, the energy barrier was 4.3 kJ/mol (Supplementary Fig. S9-C). In the gRNA-bound form of FnCas9, the lowest energy state had 0 kJ/mol energy, and other low energy states had 0.79 kJ/-mol of energy. Their inter-conversion had an energy barrier of 3.96 kJ/mol (State A to State B) and 3.17 kJ/mol (State B to State A) (Supplementary Fig. S9-D). Here, energy barriers for state transitions was lower in the gRNA bound of FnCas9 form than in its free form (Supplementary S8), suggesting switch between low energy states is easier in gRNA bound form than in the Apo form of FnCas9.

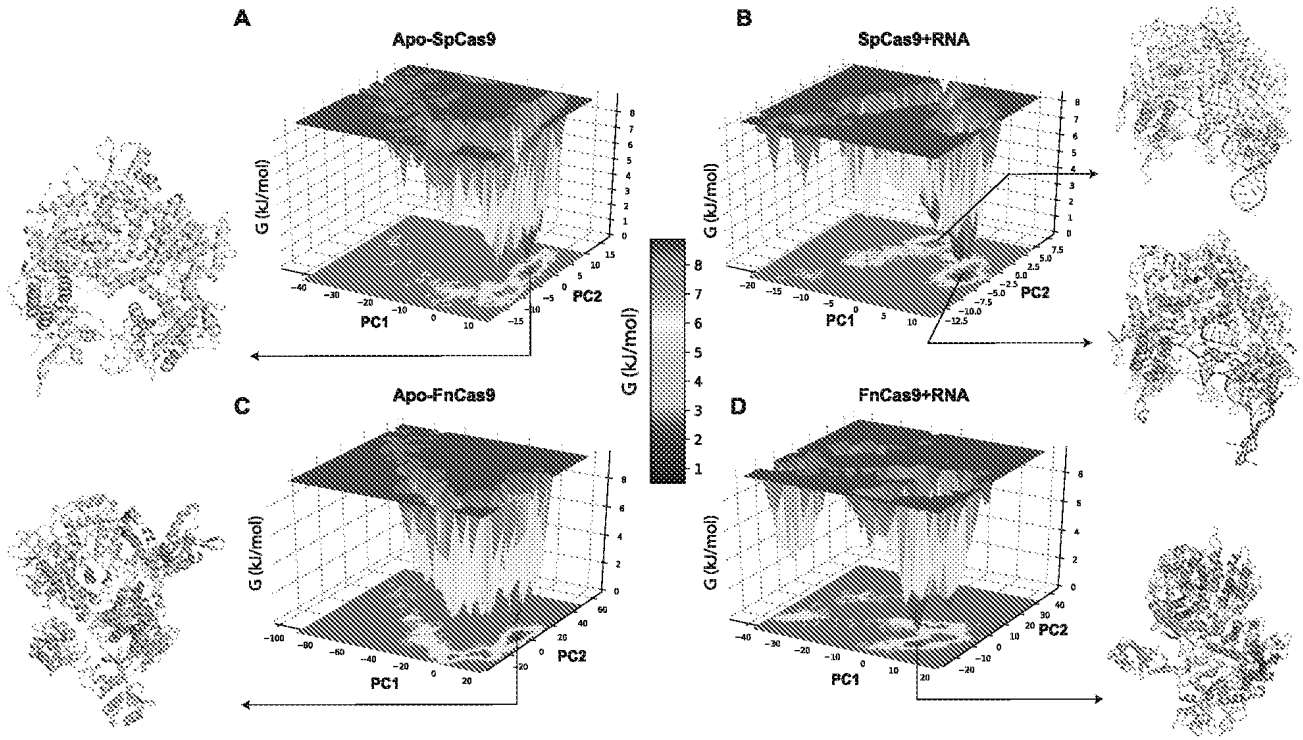


Fig. 6. A. Free energy landscape generated using principal components (PC1 and PC2) for Apo-SpCas9 and SpCas9 with gRNA during 1000 ns simulations. (B) Free energy landscape generated using principal components (PC1 and PC2) for Apo-FnCas9 and FnCas9 with gRNA during 1000 ns simulations. Energy distribution is shown by the coloring pattern: Blue defines the conformational space with minimum energy (stable state) while red defines a conformational space with maximum energy (unstable state). Principal components are displayed as a contour map at the bottom of each FEL plot with similar color pattern like the energy landscape.

2.5. Charge potentials reveal key insights into the greater specificity of FnCas9 protein

MM-PBSA/GBSA (Molecular mechanics/Poisson–Boltzmann (Generalized-Born)) is an end-state post-processing method to estimate binding free energies. This method has been proven to balance accuracy and computational efficiency, especially when dealing with large systems. In order to calculate the binding affinity of gRNA with SpCas9 and FnCas9 forms and to understand the role of electrostatics and the atomic level interactions of gRNA to the Cas9 protein, MM-GBSA calculations were performed over an ensemble of 700 snapshots extracted from 50 ns (stable short region of the MD trajectory (950–1000 ns) using gmx_MMPBSA. Per-residue energy decomposition of MM-GBSA was used to retrieve the individual amino acid contribution to the binding energy of gRNA with Cas9 protein . The average of each energy component that contributes to the binding free energy of gRNA with SpCas9 and FnCas9 is shown in Table 7A. The guide RNA (SpCas9: 85-mer, FnCas9: 94-mer) had a lesser binding affinity with SpCas9 form (−655.4 kcal/mol) than with FnCas9 form (−1254.1 kcal/mol). In both cases, the gas phase energy, especially the electrostatic energy, had contributed predominantly to the binding energy of gRNA and protein. In the case of SpCas9 (Table 7A(i)), the contribution of electrostatic energy (−2377.7 kcal/mol) was less than the polar solvation energy (2452.0 kcal/mol). On the contrary, in FnCas9 system (Table 7A (ii)), the electrostatic interaction (−4880.2 kcal/mol) had a greater contribution than the polar solvation energy (4748.2 kcal/mol). The energetic contribution of each residue was explored to identify the hotspot residues involved in the binding of Cas9 protein with gRNA (Supplementary Fig. S9). This analysis revealed that residues from the arginine-helix and REC1 domain play a pivotal role in the binding of gRNA with SpCas9 protein (Supplementary Fig. S10-A (i)); however, only a few residues had shown a greater contribution

A

Binding Energy (Average kcal/mol)

Model System	VDW	EEL	EGB	ESURF	ΔG gas	ΔG solv	ΔG Total
SpCas9 - gRNA	-649.69 +/- 61.43	-2377.69 +/- 23.96	2451.96 +/- 24.79	-79.98 +/- 1.45	-3027.39 +/- 65.44	2371.98 +/- 24.53	-651.40 +/- 61.10
FnCas9 - gRNA	-991.47 +/- 16.82	-4880.24 +/- 34.78	4748.24 +/- 33.60	-130.64 +/- 1.85	-5871.67 +/- 38.29	4617.59 +/- 33.84	-1254.07 +/- 17.07

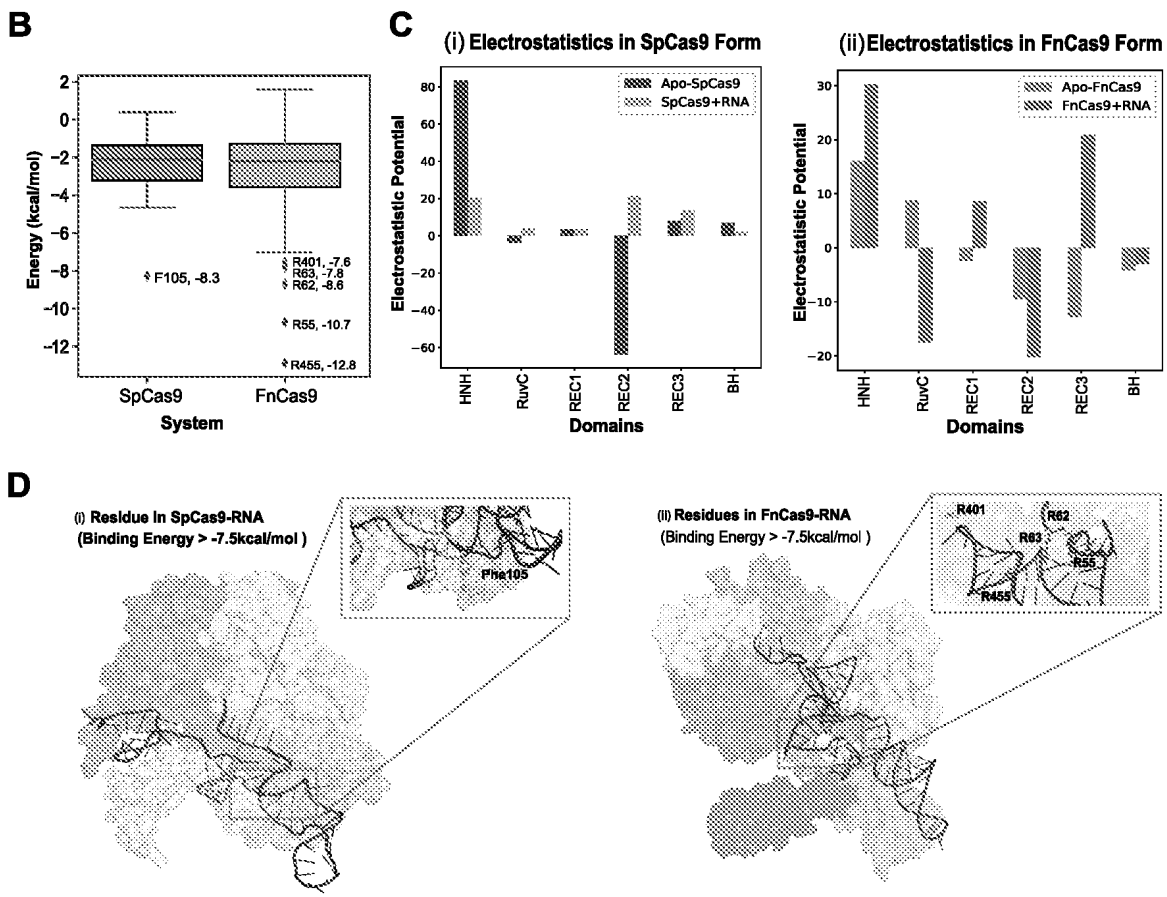


Fig. 7. A. Bar plots indicating the free energy components (vdW, electrostatic, polar surface, SASA, obtained from MM-PBSA calculations for gRNA binding with SpCas9 (left) and FnCas9 (right). B. Box plot showing the distribution of per-residue energy contribution in the binding of gRNA with SpCas9 and FnCas9. C. Electrostatic potential of HNH, RuvC, REC domains, and bridge-helix in SpCas9 and FnCas9. D. Hotspot residues involved in gRNA binding shown in sticks representation in SpCas9 and FnCas9.

to the binding energy (less than −6 kcal/mol). Many residues were found to be involved in the binding of gRNA (171) with FnCas9 protein; only those residues which have contributed more than −2 kcal/mol (<−2 kcal/mol) were plotted in the Supplementary Fig. S10-A(ii). These residues belonged to WED, REC3, REC1 domains, and the bridge-helix region of the protein. The contribution made by nucleotides of gRNA in binding energy was also shown in Supplementary Fig. S10-B. The energy decomposition distribution for both Cas9 forms was compared in Fig. 7B. The boxplot of residue-wise energy decomposition revealed that both the Cas9 forms had nearly equal median energy values with a slightly greater inter-quartile energy range for FnCas9. Interestingly, there was one outlier, Phe105 (−8.3 kcal/mol) in SpCas9 and five outliers in FnCas9, namely, Arg401 (−7.6 kcal/mol), Arg63 (−7.8 kcal/mol), Arg62 (−8.6 kcal/mol), Arg55 (−10.7 kcal/mol), and Arg455 (−12.8 kcal/mol) with energy contribution greater than −7.5 kcal/mol. These residues were highlighted on the surface of SpCas9 and FnCas9 structures in sticks representation in Fig. 7D. In FnCas9, all the outlier residues (energy < −7.5 kcal/mol) were arginines, three of which were a part of the bridge helix suggesting the pivotal role of arginine helix in gRNA recognition and binding.

Due to the strong negative electrostatic field associated with the gRNA phosphate backbone, charged interactions are expected to be critical for correct protein–RNA binding and functioning. The MMGBSA results emphasized the significance of positively charged arginine residues located in close proximity to the gRNA-binding site. This observation can be correlated with the role of electrostatic interactions in gRNA binding. To elucidate the contribution made by each Cas9 domain in maintaining a suitable environment for gRNA binding, the electrostatic potential of each domain was evaluated using the APBS [26] tool and compared with their gRNA-free and gRNA-bound forms in Fig. 8C. A highly-positive electrostatic potential was shown by the HNH domain, and a highly-negative electrostatic potential was shown by the REC2 domain in the case of Apo-SpCas9. A drastic change in electrostatic potential was observed in the case of the REC2 domain of SpCas9 after gRNA binding, where the negative potential transformed to a positive potential. In the case of FnCas9, RuvC domain transformed its potential from electropositive to electronegative. On comparing the electrostatics of domain centers after gRNA binding, it was found that FnCas9 domains had high electrostatic potential as compared to SpCas9 (Positive potential in HNH, REC1, and REC3 domain and Negative potential in RuvC and REC2 domains), supporting the observation made in work by Acharya et al.[13].

Computational Alanine Scanning is an effective method to determine the significance of side-chain functional groups of hotspot residues in protein-gRNA binding by substituting it to Alanine and then evaluating the change in binding affinity between Cas9 protein and gRNA in both the systems. The binding free-energy change due to Alanine mutation of hotspot residues (Spcas9 System: Phe105; FnCas9 System: Arg401, Arg63, Arg62, Arg55, and Arg455) was performed to assess their significance in gRNA binding. The $\Delta\Delta G_{(CAS)}$ for these residues is shown in Table 3. The alanine conversion of Phe105 decreased the binding affinity of gRNA with SpCas9 by 9.958 kcal/mol, and Arg455 decreased the binding affinity of gRNA with FnCas9 by 17 kcal/mol (similar values were observed for other hotspot residues), indicating an unfavorable alanine-substitution demonstrating the significance of these residues in gRNA binding.

3. Conclusions

Comparative analyses of gRNA-free and gRNA-bound forms of SpCas9 and FnCas9 suggested that the binding of gRNA has brought structural stability to the system. gRNA-binding has led to a decrease in structural compactness, indicating the opening of the structure and exposure of hydrophobic residues (3). The RMSF-profile for both states of SpCas9 had similar flexibility. Interestingly, in FnCas9, the Apo-form was more flexible than the gRNA-bound form (2). Residues in the REC2-REC3 domains (residues 218–233, 727–791) and the HNH (residues 950–998; 1004–1065) domain were highly fluctuating in the gRNA-free form, whereas the PI domain (residues 1565–1572) had high-mobility in the gRNA bound form of FnCas9. The FnCas9 was more volatile than the SpCas9 form, shown by the high radius of gyration, RMSD, and RMSF values. Domains with higher fluctuations had shown larger domain movement, suggesting a correlation between them (especially in Apo-FnCas9). As a consequence of gRNA-binding, a concerted series of domain movements occurred in SpCas9 (HNH-REC2, CTD-Topo, and REC3-RuvC) and FnCas9 (HNH-REC3-REC1 and PI-WED-RuvC). The FnCas9 form had shown larger domain deviations as compared to SpCas9 as their difference in average domain distance range was higher than the latter. The increase in the percentage of dynamic residues observed in both Cas9s post-gRNA-binding implies the occurrence of higher structural deformations in those regions (Supplementary Figs. 5 and 7). In SpCas9, the REC2 domain contained a higher percentage of dynamic residues, and HNH had low dynamic residue percentage. This observation was in line with conservation analysis, where the HNH domain was most conserved, and the REC2 domain was least conserved. The high fluctuations and movement shown by the PI domain in the gRNA-bound form of FnCas9 could be due to higher structural transitions and increased coil percentage after gRNA-binding. In the free-energy landscape, clusters with frames of the initial part of the trajectory were structurally and energetically distinct. Energetically more stable states were observed in the gRNA-bound form in both the Cas9s. The binding free energy analysis revealed that gRNA is more tightly bound in FnCas9 form than in SpCas9. In SpCas9, the polar-solvation energy (EGB) had contributed predominantly to the binding energy, whereas in FnCas9 electrostatic energy (EEL) was the major contributor, disclosing the significance of electrostatic interactions in gRNA-FnCas9 binding. The per-residue energy decomposition in SpCas9 form revealed the significance of arginine-helix and REC1 domain in gRNA binding. Many residues have shown contribution towards the binding of gRNA in FnCas9, mainly belonging to WED, REC3, REC1 domains, and the bridge-helix region (Supplementary Fig. S9). Hotspot-residues (energy < −7.5 kcal/mol) with the greatest contribution towards gRNA binding were Phe105 in SpCas9 and Arg455, Arg401, Arg55, Arg62, and Arg63 in FnCas9. All hotspot residues were arginines in the case of FnCas9, three of which were a part of the bridge helix (R55, R62, R63), suggesting its pivotal role in gRNA recognition and binding.

This study talks about the sequential differences as well as changes in dynamics at theatomic level between Cas9 orthologs, SpCas9 and FnCas9. It sheds light on the role of bridge-helix residues and electrostatic interactions in FnCas9-gRNA binding. Cas9 domains play an essential role in bringing the protein into the catalytic–competent stage and, ultimately, in the DNA cleavage process. The structural changes and differences in domain rearrangement in both Cas9 orthologs show that the cleavage mechanism differs in these systems, which might explain why Cas9s have different specificity and sensitivity. To gain a better understanding about the cleavage mechanism at atomic level DNA bound ternary complex of SpCas9 and FnCas9 should be simulated and compared. Future research should focus on comparing the domain rearrangement and change in binding free energies in the presence of DNA and implementing high-level quantum–mechanical simulations to describe the formation and breaking of bondsto understand the DNA cleavage mechanism of FnCas9 systems which will aid in optimizing Cas9's activity.

4. Materials and methods

4.1. Estimation of evolutionary conservation of Cas9 domains

The amino-acid positions which evolve slowly are commonly conserved sites that are important for protein structure and function. Proteins alter their secondary structure due to a change of surroundings or as a result of interaction with other proteins, but these conserved residues preserve their structures in all the conformations of the protein and therefore contribute less to the structural flexibility of the protein [27]. The ConSurf server [19] was used to estimate the level of conservation of amino acid positions in different domains of the SpCas9 and FnCas9 protein based on phylogenetic relationships between their homologous sequences. The protein structure of SpCas9 (PDB ID:4CMQ) and FnCas9 (PDB ID:5B2Q) were submitted to the ConSurf server, which calculated the conservation scores partitioned into the discrete scale of nine bins. The position with bins $>$ 8 scores is considered conserved sites, while the amino-acid positions with bin <= 2 are considered variable sites. Applying these cut-offs for indicating conserved and variable residues, the percentages of conserved and variable residues were calculated for HNH, RuvC, BH, REC1, REC2, and REC3 domains in both the proteins. The proportion of charged (H, K, R, D, E), uncharged (C, S, T, Y, Q), and non-polar (A, G, I, L, M, W, F, P, V) residues in conservation was also calculated.

4.2. System setup and MD Simulations

MD simulations have been performed using four model systems of Cas9, which were the Apo form of SpCas9 and in complex with gRNA, and Apo form of FnCas9 and in complex with gRNA. These model systems have been prepared using the crystallographic coordinates of *Streptococcus pyogenes* Apo-SpCas9 (PDB ID:4CMQ) [28], SpCas9 + RNA (PDB ID:4ZT0) [16], *Francsella novicida* Apo-FnCas9 (PDB ID:5B2Q w/o nucleic acids) [14] and FnCas9 + RNA (5B2Q, DNA strand removed), solved at 3.09 Å, 2.90 Å, and 1.70 Å resolution, respectively. The four independent MD simulations were performed using GROMACS 2020 software package [29]. AMBERff99SB and AMBER94 forcefields were used for protein, and nucleic acid parameterization [30]. The system was neutralized and solvated by explicit water molecules, which were modeled by the TIP3P) parameter set [31] in a dodecahedron box. Particle mesh Ewald was used to treat long-range electrostatic interactions [32]. The pressure and temperature were controlled with the Parrinello-Rahman coupling algorithm, and Langevin thermostat [33]. The energy minimization of the whole system was done using the steepest descent algorithm until the maximum force was less than 1000 kJ mol^{-1} nm^{-1} to remove any steric clashes. The system was heated to 300 K and equilibrated for 100 ps in both Isothermal-Isobaric and canonical ensembles. The positions of Cas9 protein and gRNA were restrained with a harmonic constant of 0.1 kcal/mol $Å^{-2}$ to avoid improper geometry. A non-restraint MD simulation was performed for 1000 ns at constant 1 atm pressure. The cut-off value for nonbonded interactions was set to 12 Å. The MD trajectory snapshot was saved after every 1 ns; thus, 1000 frames for every simulation were collected for further analysis.

4.3. Distance between domains

The center of mass of HNH, RuvC, REC1, REC2, textcolorredand REC3 domains in both the Cas9 forms was calculated using the pdb_centermass script of pdbtools (https://github.com/harmslab/pdbtools). Additionally, the COM of Topo and CTD in SpCas9 and PI and WED domains in FnCas9 protein were also determined using the same script. To quantify and compare the domain movement due to gRNA binding, the distance between the center of mass of every domain pair during the simulation time (200–1000 ns, stable part of trajectory) was calculated to characterize relative movements of intramolecular domains (due to gRNA binding) as a function of simulation time. Interdomain distances in both free and gRNA-bound forms for both the systems were averaged. The difference (Dist-Free form – Dist-bound form) in these averaged distances was determined for every domain pair. A positive delta indicates that the domains came closer, and vice versa. Normalized delta distances were taken as input to form a clustermap to illustrate the relative movement of domains due to the binding of gRNA. A clustermap uses hierarchical agglomeration clustering to cluster similar columns together. The order of clustering forms a dendrogram. These values were standard scaled to 1 to make a comparison between both Cas9s.

4.4. Domain structural transitions and fluctuations

The propensity of a residue to attain a particular secondary structural element was computed for 1000 frames (Fig. S4). A secondary structure was assigned to those residues which had a greater than 70% probability to exist either as a helix, turn, strand, or coil throughout the simulation. Those residues that did not conform to any particular secondary structure throughout our simulation (existed as a secondary structure for <70% of times throughout simulation) were also quantified and referred textcolorredto as dynamic residues. The percentage of these dynamical residues that make up each of the functional domains of the protein. To determine regions with high flexibility, the percentage of residues existing in coil conformation in more than 70% of the simulation time were calculated for every domain in Apo and gRNA-bound form for both the systems. Domain-wise atomic fluctuations were studied by plotting root-mean-square fluctuation for every domain. The average RMSF of SpCas9 and FnCas9 proteins in free and bound form were plotted.

4.5. Principal component analysis

Principal component analysis (PCA) is a useful method for assessing trajectory data from MD simulations and identifying the essential dynamics [34,35]. Based on the calculation of the covariance matrix, this method applies a linear transformation to project the original high-dimensional representation of macromolecular dynamics into the low-dimensional space. The covariance matrix of the protein backbone atoms was built and diagonalized using gmx covar method of GROMACS 5.1 package [29] to investigate the dominant motions of free and gRNA bound forms of Cas9. Only a very small number of eigenvectors (modes of fluctuation) contribute significantly to the overall motion of the protein. This can be seen from the eigenvalues contained in eigenval.xvg. The first principle component (PC1) corresponds to the system's largest amplitude motion, and the system dynamics along the PC1 is generally called "Essential Dynamics". The most dominant mode, containing 100 conformations along eigenvector 1 was saved in PDB format and visualized in Normal Mode Wizard of VMD [25].

4.6. MMGBSA calculations

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method [36] is a popular technique for determining binding free energies between proteins and ligands. gmx_MMPBSA [37] was used to calculate the enthalpy contribution to binding free energy. The GB (OBC) model (igb = 2) was used to calculate the solvation energy. A total of 700 snapshots, equally distributed in the

last 50 ns of the SpCas9 + RNA and FnCas9 + RNA trajectory were selected to calculate the enthalpy contribution using MM/GBSA. The single trajectory approach was used to calculate the binding free energy. Molecular mechanics potential energy (electrostatic and van der Waals energies), the free energy of solvation (polar and nonpolar solvation energies), and the energetic contribution of each residue and nucleotide to total binding energy were obtained for both SpCas9 and FnCas9 proteins with gRNA. The gas-phase energies calculated using the Molecular Mechanics force field, are a sum of van der Waals (VDWAALS) and electrostatic (EEL) energies, whereas the solvation energy includes a polar and a non-polar component. The polar component is calculated using the implicit Solvent Model, i.e. Generalized Born (GB), (EGB) and the non-Polar component depends on solvent accessible surface area (ESURF). Free energies of solvation are estimated by applying Poisson–Boltzmann (PB) calculations for the electrostatic contribution and a surface-area-dependent term for the non-electrostatic contribution to solvation.

4.7. Electrostatic calculation of Cas9 domains

The electrostatic contribution made by HNH, RuvC, REC1, REC2, and REC3 domains were calculated using the Adaptive Poisson–Boltzmann Solver (APBS) [38]. Prior to using APBS, the standalone version of PDB2PQR [39] was used to assign charge and radius parameters for the CHARMM force field to all the domain structures. APBS uses iterative solvers to solve the nonlinear algebraic equations resulting from the discretized Poisson–Boltzmann equation with a fixed error tolerance of 10÷6. The electrostatic values for the centroid coordinates of each of these domains were then extracted from its corresponding electrostatic potential file using the multivalue script (provided in APBS).

4.8. Gibbs free energy landscape

The free energy landscape (FEL) is used for describing conformation exchange in biomolecular processes such as molecular recognition, folding, and aggregation. The population and stability of functionally different states are determined by free energy basins and their depths, while the inter-basin barriers correspond to the transient states that connect them. gmx sham module of GROMACS was used to compute Gibbs Free Energy using the Boltzmann equation. The relative free energy between two states is given by:

$$\Delta G = -KbT * ln[P1(X)/P2(X)]$$

where Kb is the Boltzmann constant, T is absolute temperature, X stands for the reaction coordinate and P(X) is the probability distribution of the system along the reaction coordinate. Based on the PCA calculation, PC1 and PC2 were chosen as reaction coordinates in this analysis. Frame-wise free energy profiles of all four Cas9s are shown in Supplementary Fig. S9.

4.9. Computational alanine scanning

Computational Alanine Scanning is an energy-based method to calculate the effects of alanine mutations on the binding free energy of a protein-gRNA complex. This alanine substitution removes the side chain from the selected residue which aids in the assessment of the role of a side chain functional groups at a particular position. It is used to determine the significance of a residue in protein-gRNA binding. The principle followed is:

$$\Delta\Delta G^{(CAS)} = \Delta Gbind^{(Ala)} - \Delta Gbind^{(Wild-type)}$$

In this formula, $\Delta\Delta G_{(CAS)}$ is the binding free energy difference between $\Delta Gbind_{(Ala)}$ (residue substituted with Alanine) and $\Delta Gbind_{(Wild-type)}$. This analysis was performed to evaluate the potential of hotspot residues in gRNA binding. 50 frames from the last 10 ns part of the trajectory were considered as input for both the systems.

Author contribution

G.P. was involved in acquisition of data, analysis and/or interpretation of data, drafting the manuscript and revising the manuscript critically for important intellectual content. A.R. led the conception and design of study, analysis and interpretation of the data, preparing the final manuscript and its revision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Dr. Arul Murugan (Indraprastha Institute of Information Technology, Delhi) for his critical insights and valuable suggestions in manuscript writing. We would also like to thank Dr. Debajyoti Chakraborty (CSIR-Institute of Genomics and Integrative Biology) for reviewing the manuscript. The authors would like to thank the computational facility at IIIT Delhi for supporting this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.csbj.2022.07.041.

References

[1] Brouns Stan JJ, Jore Matthijs M, Lundgren Magnus, Westra Edze R, Slijkhuis Rik JH, Snijders Ambrosius PL, Dickman Mark J, Makarova Kira S, Koonin Eugene V, Oost John van der, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science (New York, N.Y.) 2008;321(5891):960–4. https://doi.org/10.1126/science.1159689. ISSN 1095–9203 (Electronic).

[2] Barrangou Rodolphe, Fremaux Christophe, Deveau Hélène, Richards Melissa, Boyaval Patrick, Moineau Sylvain, Romero Dennis A, Horvath Philippe, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science (New York, N.Y.) 2007;315(5819):1709–12. https://doi.org/10.1126/science.1138140. ISSN 1095–9203 (Electronic).

[3] Chen Hongfan, Choi Jihoon, Bailey Scott. Cut Site Selection by the Two Nuclease Domains of the Cas9 RNA-guided Endonuclease. J Biol Chem 2014;289 (19):13284–94. https://doi.org/10.1074/jbc.M113.539726. ISSN 0021–9258.

[4] Nishimasu Hiroshi, Ann Ran F, Hsu Patrick D, Konermann Silvana, Shehata Soraya I, Dohmae Naoshi, Ishitani Ryuichiro, Zhang Feng, Nureki Osamu, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014;156(5):935–49. https://doi.org/10.1016/j.cell.2014.02.001. ISSN 1097–4172 (Electronic).

[5] Josephs Eric A, Dewran Kocak D, Fitzgibbon Christopher J, McMenemy Joshua, Gersbach Charles A, Marszalek Piotr E. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. Nucl Acids Res 2015;43(18):8924–41. https://doi.org/10.1093/nar/gkv893. ISSN 1362–4962 (Electronic).

[6] Zhang Xiao-Hui, Tee Louis Y, Wang Xiao-Gang, Huang Qun-Shan, Yang Shi-Hua. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular therapy. Nucl Acids 2015;4(11):e264. https://doi.org/10.1038/mtna.2015.37. ISSN 2162–2531 (Print).

[7] Cho Seung Woo, Kim Sojung, Kim Yongsub, Kweon Jiyeon, Kim Heon Seok, Bae Sangsu, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res 2014;24(1):132–41. https://doi.org/10.1101/gr.162339.113. ISSN 1549–5469 (Electronic).

[8] Wu Xuebing, Kriz Andrea J, Sharp Phillip A. Target specificity of the CRISPR-Cas9 system. Quantitative Biol (Beijing, China) 2014;2(2):59–70. https://doi.org/10.1007/s40484-014-0030-x. ISSN 2095–4689 (Print).

[9] Manghwar Hakim, Lindsey Keith, Zhang Xianlong, Jin Shuangxia. CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. Trends Plant Sci 2019;24(12):1102–25. https://doi.org/10.1016/j.tplants.2019.09.006. ISSN 1360–1385.

[10] Casini Antonio, Olivieri Michele, Petris Gianluca, Montagna Claudia, Reginato Giordano, Maule Giulia, Lorenzin Francesca, Prandi Davide, Romanel Alessandro, Demichelis Francesca, Inga Alberto, Cereseto Anna, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nat Biotechnol 2018;36(3):265–71. https://doi.org/10.1038/nbt.4066. ISSN 1546–1696.

[11] Lee Jungjoon K, Jeong Euihwan, Lee Joonsun, Jung Minhee, Shin Eunji, Kim Young-hoon, Lee Kangin, Jung Inyoung, Kim Daesik, Kim Seokjoong, Kim Jin-Soo, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nat Commun 2018;9(1):3048. https://doi.org/10.1038/s41467-018-05477-x. ISSN 2041–1723.

[12] Vakulskas Christopher A, Dever Daniel P, Rettig Garrett R, Turk Rolf, Jacobi Ashley M, Collingwood Michael A, Bode Nicole M, McNeill Matthew S, Yan Shuqi, Camarena Joab, Lee Ciaran M, Park So Hyun, Wiebking Volker, Bak Rasmus O, Gomez-Ospina Natalia, Pavel-Dinu Mara, Sun Wenchao, Bao Gang, Porteus Matthew H, Behlke Mark A, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. Nat Med 2018;24 (8):1216–24. https://doi.org/10.1038/s41591-018-0137-0. ISSN 1546–170X.

[13] Acharya Sundaram, Mishra Arpit, Paul Deepanjan, Ansari Asgar Hussain, Azhar Mohd, Kumar Manoj, Rauthan Riya, Sharma Namrata, Aich Meghali, Sinha Dipanjali, Sharma Saumya, Jain Shivani, Ray Arjun, Jain Suman, Ramalingam Sivaprakash, Maiti Souvik, Chakraborty Debojyoti, et al. Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. Proc Natl Acad Sci 2019;116(42):20959–68. https://doi.org/10.1073/pnas.1818461116. ISSN 0027–8424. https://www.pnas.org/content/116/42/20959.

[14] Hirano Hisato, Gootenberg Jonathan S, Horii Takuro, Abudayyeh Omar O, Kimura Mika, Hsu Patrick D, Nakane Takanori, Ishitani Ryuichiro, Hatada Izuho, Zhang Feng, Nishimasu Hiroshi, Nureki Osamu, et al. Structure and Engineering of Francisella novicida Cas9. Cell 2016;164(5):950–61. https://doi.org/10.1016/j.cell.2016.01.039. ISSN 1097–4172 (Electronic).

[15] Palermo Giulia, Miao Yinglong, Walker Ross C, Jinek Martin, Andrew McCammon J, et al. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. ACS Central Sci 2016;2 (10):756–63. https://doi.org/10.1021/acscentsci.6b00318. ISSN 2374–7943.

[16] Jiang Fuguo, Zhou Kaihong, Ma Linlin, Gressel Saskia, Doudna Jennifer A. STRUCTURAL BIOLOGY.A Cas9-guide RNA complex preorganized for target DNA recognition. Science (New York, N.Y.) 2015;348(6242):1477–81. ISSN 1095–9203 (Electronic).

[17] Palermo Giulia, Miao Yinglong, Walker Ross C, Jinek Martin, Andrew McCammon J. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proc Natl Acad Sci 2017;114(28):7260–5. https://doi.org/10.1073/pnas.1707645114.

[18] Palermo Giulia. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. J Chem Inf Model 2019;59(5):2394–406. https://doi.org/10.1021/acs.jcim.8b00888. ISSN 1549–9596.

[19] Ashkenazy Haim, Abadi Shiran, Martz Eric, Chay Ofer, Mayrose Itay, Pupko Tal, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl Acids Res 2016;44(W1): W344–50. https://doi.org/10.1093/nar/gkw408. ISSN 0305–1048.

[20] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16(6):276–7. https://doi.org/10.1016/s0168-9525(00)02024-2. ISSN 0168–9525 (Print).

[21] Jiang Fuguo, Doudna Jennifer A. CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys 2017;46(1):505–29. https://doi.org/10.1146/annurev-biophys-062215-010822.

[22] Sternberg Samuel H, LaFrance Benjamin, Kaplan Matias, Doudna Jennifer A. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 2015;527(7576):110–3. https://doi.org/10.1038/nature15544. ISSN 1476–4687 (Electronic).

[23] Palermo Giulia, Chen Janice S, Ricci Clarisse G, Rivalta Ivan, Jinek Martin, Batista Victor S, Doudna Jennifer A, Andrew J. McCammon: Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. Q Rev Biophys 2018. https://doi.org/10.1017/S0033583518000070. ISSN 1469–8994 (Electronic).

[24] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–27. ISSN 0887-3585 (Print). doi: 10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7.

[25] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graphics 1996;14(1). https://doi.org/10.1016/0263-7855(96)00018-5. pp. 27–28,33–38. ISSN 0263–7855 (Print).

[26] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98(18):10037–41. https://doi.org/10.1073/pnas.181342398. ISSN 0027–8424 (Print).

[27] Chen Ke, Kurgan Lukasz A, Ruan Jishou. Prediction of flexible/rigid regions from protein sequences using k-spaced amino acid pairs. BMC Struct Biol 2007;7(1):25. https://doi.org/10.1186/1472-6807-7-25. ISSN 1472–6807.

[28] Jinek Martin, Jiang Fuguo, Taylor David W, Sternberg Samuel H, Kaya Emine, Ma Enbo, Anders Carolin, Hauer Michael, Zhou Kaihong, Lin Steven, Kaplan Matias, Iavarone Anthony T, Charpentier Emmanuelle, Nogales Eva, Doudna Jennifer A, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 2014;343(6176):1247997. https://doi.org/10.1126/science.1247997. ISSN 1095–9203 (Electronic).

[29] Abraham Mark James, Murtola Teemu, Schulz Roland, Páll Szilárd, Smith Jeremy C, Hess Berk, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001. ISSN 2352–7110.

[30] Guy Andrew T, Piggot Thomas J, Khalid Syma. Single-Stranded DNA within Nanopores: Conformational Dynamics and Implications for Sequencing; a Molecular Dynamics Simulation Study. Biophys J 2012;103(5):1028–36. https://doi.org/10.1016/j.bpj.2012.08.012. ISSN 0006–3495.

[31] Jorgensen William L, Chandrasekhar Jayaraman, Madura Jeffry D, Impey Roger W, Klein Michael L. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79(2):926–35. https://doi.org/10.1063/1.445869. ISSN 0021–9606.

[32] Darden Tom, York Darrin, Pedersen Lee. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J Chem Phys 1993;98(12):10089–92. https://doi.org/10.1063/1.464397. ISSN 0021–9606.

[33] Hatano Takahiro, Sasa Shin-ichi. Steady-State Thermodynamics of Langevin Systems. Phys Rev Lett 2001;86(16):3463–6. https://doi.org/10.1103/PhysRevLett.86.3463.

[34] Wan Hua, Hu Jian-ping, Tian Xu-hong, Chang Shan. Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d. Phys Chem Chem Phys 2013;15(4):1241–51. https://doi.org/10.1039/C2CP43830D.

[35] Wan Hua, Chang Shan, Hu Jian-ping, Tian Yuan-xin, Tian Xu-hong. Molecular Dynamics Simulations of Ternary Complexes: Comparisons of LEAFY Protein Binding to Different DNA Motifs. J Chem Inf Model 2015;55(4):784–94. https://doi.org/10.1021/ci5006705i. ISSN 1549–960X (Electronic).

[36] Genheden Samuel, Ryde Ulf. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discovery 2015;10 (5):449–61. https://doi.org/10.1517/17460441.2015.1032936. ISSN 1746–045X (Electronic).

[37] Valdés-Tresanco Mario S, Valdés-Tresanco Mario E, Valiente Pedro A, Moreno Ernesto. gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. J Chem Theory Comput 2021;17(10):6281–91. https://doi.org/10.1021/acs.jctc.1c00645. ISSN 1549–9618.

[38] Jurrus Elizabeth, Engel Dave, Star Keith, Monson Kyle, Brandi Juan, Felberg Lisa E, Brookes David H, Wilson Leighton, Chen Jiahui, Liles Karina, Chun Minju, Li Peter, Gohara David W, Dolinsky Todd, Konecny Robert, Koes David R, Nielsen Jens Erik, Head-Gordon Teresa, Geng Weihua, Krasny Robert, Wei Guo-Wei, Holst Michael J, Andrew McCammon J, Baker Nathan A, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci 2018;27 (1):112–28. https://doi.org/10.1002/pro.3280. ISSN 0961–8368.

[39] Dolinsky Todd J, Nielsen Jens E, Andrew McCammon J, Baker Nathan A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucl Acids Res 2004;32(Web Server issue): W665–7. https://doi.org/10.1093/nar/gkh381. ISSN 1362–4962 (Electronic).