Supplementary Information (SI) for Physical Chemistry Chemical Physics.

Machine Learning Models for Predicting Configuration of Modified Knuckle Epitope Peptides of BMP-2 Protein Using Mesoscale Simulation Data

Ricky Anshuman Dash[1] and Esmaiel Jabbari[1]*

[1]Biomimetic Materials and Tissue Engineering Laboratory, Chemical Engineering Department,

University of South Carolina, 301 Main Street, Columbia, SC, USA, 29208

Submitted to

RSC Physical Chemistry Chemical Physics

***Corresponding author:**
Esmaiel Jabbari, Ph.D.
Full Professor of Chemical and Biomedical Engineering
Department of Chemical Engineering
Swearingen Engineering Center, Rm 3C13
University of South Carolina
Columbia, SC 29208
Tel: (803) 777-8022
Fax: (803) 777-0973
E-mail: jabbari@cec.sc.edu

SUPPLEMENTARY TABLE 1: List of the modified BMP-2 KEP sequences.

KIPKASSVPTELSAISTLYL	KIPKAYSIPTELSAIYTLYL	KINKASSVPTINSAISTLYL	KIPKASSVETELSEISTLHH
KIPKASSFPTELSAISTLYL	KIPKAYSFPTELSAIITLYL	KIVKASSVPTFNSAISTLYL	KIPKASSVETELSEISTLHR
KIPKASSPPTELSAISTLYL	KIPKAYSPPTELSAIITLYL	KIVKASSVPTLNSAISTLYL	KIPKASSVETELSEISTLRD
KIPKASSTPTELSAISTLYL	KIPKAYSTPTELSAIITLYL	KIVKASSVPTTNSAISTLYL	KIPKASSVETELSEISTLRE
KIPKASSSPTELSAISTLYL	KIPKAYSSPTELSAIITLYL	KIVKASSVPTNNSAISTLYL	KIPKASSVETELSEISTLRH
KIPKASSYPTELSAISTLYL	KIPKAYSYPTELSAIITLYL	KIVKASSVPTVNSAISTLYL	KIPKASSVETELSEISTLRR
KIPKASSIPTELSAISTLYL	KIPKAYSIPTELSAIITLYL	KIVKASSVPTINSAISTLYL	KIPKASSVETELSHISTLDD
KIPKASSIPTELSAIITLYL	KIPKAISFPTELSAIFTLYL	KIIKASSVPTFNSAISTLYL	KIPKASSVETELSHISTLDE
KIPKASSFPTELSAIITLYL	KIPKAISPPTELSAIFTLYL	KIIKASSVPTLNSAISTLYL	KIPKASSVETELSHISTLDH
KIPKASSPPTELSAIITLYL	KIPKAISTPTELSAIFTLYL	KIIKASSVPTTNSAISTLYL	KIPKASSVETELSHISTLDR
KIPKASSTPTELSAIITLYL	KIPKAISSPTELSAIFTLYL	KIIKASSVPTNNSAISTLYL	KIPKASSVETELSHISTLED
KIPKASSSPTELSAIITLYL	KIPKAISYPTELSAIFTLYL	KIIKASSVPTVNSAISTLYL	KIPKASSVETELSHISTLEE
KIPKASSYPTELSAIITLYL	KIPKAISIPTELSAIFTLYL	KIIKASSVPTINSAISTLYL	KIPKASSVETELSHISTLEH
KIPKASSFPTELSAIFTLYL	KIPKAISFPTELSAIPTLYL	KIFKASSVPTFVSAISTLYL	KIPKASSVETELSHISTLER
KIPKAFSFPTELSAIFTLYL	KIPKAISPPTELSAIPTLYL	KIFKASSVPTLVSAISTLYL	KIPKASSVETELSHISTLHD
KIPKAFSPPTELSAIFTLYL	KIPKAISTPTELSAIPTLYL	KIFKASSVPTTVSAISTLYL	KIPKASSVETELSHISTLHE
KIPKAFSTPTELSAIFTLYL	KIPKAISSPTELSAIPTLYL	KIFKASSVPTNVSAISTLYL	KIPKASSVETELSHISTLHH
KIPKAFSSPTELSAIFTLYL	KIPKAISYPTELSAIPTLYL	KIFKASSVPTVVSAISTLYL	KIPKASSVETELSHISTLHR
KIPKAFSYPTELSAIFTLYL	KIPKAISIPTELSAIPTLYL	KIFKASSVPTIVSAISTLYL	KIPKASSVETELSHISTLRD
KIPKAFSIPTELSAIFTLYL	KIPKAISFPTELSAITTLYL	KILKASSVPTFVSAISTLYL	KIPKASSVETELSHISTLRE
KIPKAFSFPTELSAIPTLYL	KIPKAISPPTELSAITTLYL	KILKASSVPTLVSAISTLYL	KIPKASSVETELSHISTLRH
KIPKAFSPPTELSAIPTLYL	KIPKAISTPTELSAITTLYL	KILKASSVPTTVSAISTLYL	KIPKASSVETELSHISTLRR
KIPKAFSTPTELSAIPTLYL	KIPKAISSPTELSAITTLYL	KILKASSVPTNVSAISTLYL	KIPKASSVETELSRISTLDD
KIPKAFSSPTELSAIPTLYL	KIPKAISYPTELSAITTLYL	KILKASSVPTVVSAISTLYL	KIPKASSVETELSRISTLDE
KIPKAFSYPTELSAIPTLYL	KIPKAISIPTELSAITTLYL	KILKASSVPTIVSAISTLYL	KIPKASSVETELSRISTLDH
KIPKAFSIPTELSAIPTLYL	KIPKAISFPTELSAISTLYL	KITKASSVPTFVSAISTLYL	KIPKASSVETELSRISTLDR
KIPKAFSFPTELSAITTLYL	KIPKAISPPTELSAISTLYL	KITKASSVPTLVSAISTLYL	KIPKASSVETELSRISTLED
KIPKAFSPPTELSAITTLYL	KIPKAISTPTELSAISTLYL	KITKASSVPTTVSAISTLYL	KIPKASSVETELSRISTLEE
KIPKAFSTPTELSAITTLYL	KIPKAISSPTELSAISTLYL	KITKASSVPTNVSAISTLYL	KIPKASSVETELSRISTLEH
KIPKAFSSPTELSAITTLYL	KIPKAISYPTELSAISTLYL	KITKASSVPTVVSAISTLYL	KIPKASSVETELSRISTLER
KIPKAFSYPTELSAITTLYL	KIPKAISIPTELSAISTLYL	KITKASSVPTIVSAISTLYL	KIPKASSVETELSRISTLHD
KIPKAFSIPTELSAITTLYL	KIPKAISFPTELSAIYTLYL	KINKASSVPTFVSAISTLYL	KIPKASSVETELSRISTLHE
KIPKAFSFPTELSAISTLYL	KIPKAISPPTELSAIYTLYL	KINKASSVPTLVSAISTLYL	KIPKASSVETELSRISTLHH
KIPKAFSPPTELSAISTLYL	KIPKAISTPTELSAIYTLYL	KINKASSVPTTVSAISTLYL	KIPKASSVETELSRISTLHR
KIPKAFSTPTELSAISTLYL	KIPKAISSPTELSAIYTLYL	KINKASSVPTNVSAISTLYL	KIPKASSVETELSRISTLRD
KIPKAFSSPTELSAISTLYL	KIPKAISYPTELSAIYTLYL	KINKASSVPTVVSAISTLYL	KIPKASSVETELSRISTLRE
KIPKAFSYPTELSAISTLYL	KIPKAISIPTELSAIYTLYL	KINKASSVPTIVSAISTLYL	KIPKASSVETELSRISTLRH
KIPKAFSIPTELSAISTLYL	KIPKAISFPTELSAIITLYL	KIVKASSVPTFVSAISTLYL	KIPKASSVETELSRISTLRR
KIPKAFSFPTELSAIYTLYL	KIPKAISIPTELSAIITLYL	KIVKASSVPTLVSAISTLYL	KIPKASSVHTELSDISTLDD
KIPKAFSPPTELSAIYTLYL	KIPKAISPPTELSAIITLYL	KIVKASSVPTTVSAISTLYL	KIPKASSVHTELSDISTLDE
KIPKAFSTPTELSAIYTLYL	KIPKAISTPTELSAIITLYL	KIVKASSVPTNVSAISTLYL	KIPKASSVHTELSDISTLDH
KIPKAFSSPTELSAIYTLYL	KIPKAISSPTELSAIITLYL	KIVKASSVPTVVSAISTLYL	KIPKASSVHTELSDISTLDR
KIPKAFSYPTELSAIYTLYL	KIPKAISYPTELSAIITLYL	KIVKASSVPTIVSAISTLYL	KIPKASSVHTELSDISTLED
KIPKAFSIPTELSAIYTLYL	KIPKASSVPTFLSAISTLYL	KIIKASSVPTFVSAISTLYL	KIPKASSVHTELSDISTLEE
KIPKAFSFPTELSAIITLYL	KIFKASSVPTFLSAISTLYL	KIIKASSVPTLVSAISTLYL	KIPKASSVHTELSDISTLEH
KIPKAFSPPTELSAIITLYL	KIFKASSVPTFFSAISTLYL	KIIKASSVPTTVSAISTLYL	KIPKASSVHTELSDISTLER
KIPKAFSTPTELSAIITLYL	KIFKASSVPTLFSAISTLYL	KIIKASSVPTNVSAISTLYL	KIPKASSVHTELSDISTLHD
KIPKAFSSPTELSAIITLYL	KIFKASSVPTTFSAISTLYL	KIIKASSVPTVVSAISTLYL	KIPKASSVHTELSDISTLHE
KIPKAFSYPTELSAIITLYL	KIFKASSVPTNFSAISTLYL	KIIKASSVPTIVSAISTLYL	KIPKASSVHTELSDISTLHH
KIPKAFSIPTELSAIITLYL	KIFKASSVPTVFSAISTLYL	KIFKASSVPTFISAISTLYL	KIPKASSVHTELSDISTLHR
KIPKAPSFPTELSAIFTLYL	KIFKASSVPTIFSAISTLYL	KIFKASSVPTLISAISTLYL	KIPKASSVHTELSDISTLRD
KIPKAPSPPTELSAIFTLYL	KILKASSVPTFFSAISTLYL	KIFKASSVPTTISAISTLYL	KIPKASSVHTELSDISTLRE
KIPKAPSTPTELSAIFTLYL	KILKASSVPTLFSAISTLYL	KIFKASSVPTNISAISTLYL	KIPKASSVHTELSDISTLRH
KIPKAPSSPTELSAIFTLYL	KILKASSVPTTFSAISTLYL	KIFKASSVPTVISAISTLYL	KIPKASSVHTELSDISTLRR

KIPKAPSYPTELSAIFTLYL	KILKASSVPTNFSAISTLYL	KIFKASSVPTIISAISTLYL	KIPKASSVHTELSEISTLDD
KIPKAPSIPTELSAIFTLYL	KILKASSVPTVFSAISTLYL	KILKASSVPTFISAISTLYL	KIPKASSVHTELSEISTLDE
KIPKAPSFPTELSAIPTLYL	KILKASSVPTIFSAISTLYL	KILKASSVPTLISAISTLYL	KIPKASSVHTELSEISTLDH
KIPKAPSPPTELSAIPTLYL	KITKASSVPTFFSAISTLYL	KILKASSVPTTISAISTLYL	KIPKASSVHTELSEISTLDR
KIPKAPSTPTELSAIPTLYL	KITKASSVPTLFSAISTLYL	KILKASSVPTNISAISTLYL	KIPKASSVHTELSEISTLED
KIPKAPSSPTELSAIPTLYL	KITKASSVPTTFSAISTLYL	KILKASSVPTVISAISTLYL	KIPKASSVHTELSEISTLEE
KIPKAPSYPTELSAIPTLYL	KITKASSVPTNFSAISTLYL	KILKASSVPTIISAISTLYL	KIPKASSVHTELSEISTLEH
KIPKAPSIPTELSAIPTLYL	KITKASSVPTVFSAISTLYL	KITKASSVPTFISAISTLYL	KIPKASSVHTELSEISTLER
KIPKAPSFPTELSAITTLYL	KITKASSVPTIFSAISTLYL	KITKASSVPTLISAISTLYL	KIPKASSVHTELSEISTLHD
KIPKAPSPPTELSAITTLYL	KINKASSVPTFFSAISTLYL	KITKASSVPTTISAISTLYL	KIPKASSVHTELSEISTLHE
KIPKAPSTPTELSAITTLYL	KINKASSVPTLFSAISTLYL	KITKASSVPTNISAISTLYL	KIPKASSVHTELSEISTLHH
KIPKAPSSPTELSAITTLYL	KINKASSVPTTFSAISTLYL	KITKASSVPTVISAISTLYL	KIPKASSVHTELSEISTLHR
KIPKAPSYPTELSAITTLYL	KINKASSVPTNFSAISTLYL	KITKASSVPTIISAISTLYL	KIPKASSVHTELSEISTLRD
KIPKAPSIPTELSAITTLYL	KINKASSVPTVFSAISTLYL	KINKASSVPTFISAISTLYL	KIPKASSVHTELSEISTLRE
KIPKAPSFPTELSAISTLYL	KINKASSVPTIFSAISTLYL	KINKASSVPTLISAISTLYL	KIPKASSVHTELSEISTLRH
KIPKAPSPPTELSAISTLYL	KIVKASSVPTFFSAISTLYL	KINKASSVPTTISAISTLYL	KIPKASSVHTELSEISTLRR
KIPKAPSTPTELSAISTLYL	KIVKASSVPTLFSAISTLYL	KINKASSVPTNISAISTLYL	KIPKASSVHTELSHISTLDD
KIPKAPSSPTELSAISTLYL	KIVKASSVPTTFSAISTLYL	KINKASSVPTVISAISTLYL	KIPKASSVHTELSHISTLDE
KIPKAPSYPTELSAISTLYL	KIVKASSVPTNFSAISTLYL	KINKASSVPTIISAISTLYL	KIPKASSVHTELSHISTLDH
KIPKAPSIPTELSAISTLYL	KIVKASSVPTVFSAISTLYL	KIVKASSVPTFISAISTLYL	KIPKASSVHTELSHISTLDR
KIPKAPSFPTELSAIYTLYL	KIVKASSVPTIFSAISTLYL	KIVKASSVPTLISAISTLYL	KIPKASSVHTELSHISTLED
KIPKAPSPPTELSAIYTLYL	KIIKASSVPTFFSAISTLYL	KIVKASSVPTTISAISTLYL	KIPKASSVHTELSHISTLEE
KIPKAPSTPTELSAIYTLYL	KIIKASSVPTLFSAISTLYL	KIVKASSVPTNISAISTLYL	KIPKASSVHTELSHISTLEH
KIPKAPSSPTELSAIYTLYL	KIIKASSVPTTFSAISTLYL	KIVKASSVPTVISAISTLYL	KIPKASSVHTELSHISTLER
KIPKAPSYPTELSAIYTLYL	KIIKASSVPTNFSAISTLYL	KIVKASSVPTIISAISTLYL	KIPKASSVHTELSHISTLHD
KIPKAPSIPTELSAIYTLYL	KIIKASSVPTVFSAISTLYL	KIIKASSVPTFISAISTLYL	KIPKASSVHTELSHISTLHE
KIPKAPSFPTELSAIITLYL	KIIKASSVPTIFSAISTLYL	KIIKASSVPTLISAISTLYL	KIPKASSVHTELSHISTLHH
KIPKAPSPPTELSAIITLYL	KIFKASSVPTLLSAISTLYL	KIIKASSVPTTISAISTLYL	KIPKASSVHTELSHISTLHR
KIPKAPSTPTELSAIITLYL	KIFKASSVPTTLSAISTLYL	KIIKASSVPTNISAISTLYL	KIPKASSVHTELSHISTLRD
KIPKAPSSPTELSAIITLYL	KIFKASSVPTNLSAISTLYL	KIIKASSVPTVISAISTLYL	KIPKASSVHTELSHISTLRE
KIPKAPSYPTELSAIITLYL	KIFKASSVPTVLSAISTLYL	KIIKASSVPTIISAISTLYL	KIPKASSVHTELSHISTLRH
KIPKAPSIPTELSAIITLYL	KIFKASSVPTILSAISTLYL	KIPKASSVDTELSDISTLDD	KIPKASSVHTELSHISTLRR
KIPKATSFPTELSAIFTLYL	KILKASSVPTFLSAISTLYL	KIPKASSVDTELSDISTLDE	KIPKASSVHTELSRISTLDD
KIPKATSPPTELSAIFTLYL	KILKASSVPTLLSAISTLYL	KIPKASSVDTELSDISTLDH	KIPKASSVHTELSRISTLDE
KIPKATSTPTELSAIFTLYL	KILKASSVPTTLSAISTLYL	KIPKASSVDTELSDISTLDR	KIPKASSVHTELSRISTLDH
KIPKATSSPTELSAIFTLYL	KILKASSVPTNLSAISTLYL	KIPKASSVDTELSDISTLED	KIPKASSVHTELSRISTLDR
KIPKATSYPTELSAIFTLYL	KILKASSVPTVLSAISTLYL	KIPKASSVDTELSDISTLEE	KIPKASSVHTELSRISTLED
KIPKASSSSSELSASSSSSS	KILKASSVPTILSAISTLYL	KIPKASSVDTELSDISTLEH	KIPKASSVHTELSRISTLEE
KIPKATSIPTELSAIFTLYL	KITKASSVPTFLSAISTLYL	KIPKASSVDTELSDISTLER	KIPKASSVHTELSRISTLEH
KIPKATSFPTELSAIPTLYL	KITKASSVPTLLSAISTLYL	KIPKASSVDTELSDISTLHD	KIPKASSVHTELSRISTLER
KIPKATSPPTELSAIPTLYL	KITKASSVPTTLSAISTLYL	KIPKASSVDTELSDISTLHE	KIPKASSVHTELSRISTLHD
KIPKATSTPTELSAIPTLYL	KITKASSVPTNLSAISTLYL	KIPKASSVDTELSDISTLHH	KIPKASSVHTELSRISTLHE
KIPKATSSPTELSAIPTLYL	KITKASSVPTVLSAISTLYL	KIPKASSVDTELSDISTLHR	KIPKASSVHTELSRISTLHH
KIPKATSYPTELSAIPTLYL	KITKASSVPTILSAISTLYL	KIPKASSVDTELSDISTLRD	KIPKASSVHTELSRISTLHR
KIPKATSIPTELSAIPTLYL	KINKASSVPTFLSAISTLYL	KIPKASSVDTELSDISTLRE	KIPKASSVHTELSRISTLRD
KIPKATSFPTELSAITTLYL	KINKASSVPTLLSAISTLYL	KIPKASSVDTELSDISTLRH	KIPKASSVHTELSRISTLRE
KIPKATSPPTELSAITTLYL	KINKASSVPTTLSAISTLYL	KIPKASSVDTELSDISTLRR	KIPKASSVHTELSRISTLRH
KIPKATSTPTELSAITTLYL	KINKASSVPTNLSAISTLYL	KIPKASSVDTELSEISTLDD	KIPKASSVHTELSRISTLRR
KIPKATSSPTELSAITTLYL	KINKASSVPTVLSAISTLYL	KIPKASSVDTELSEISTLDE	KIPKASSVRTELSDISTLDD
KIPKATSYPTELSAITTLYL	KINKASSVPTILSAISTLYL	KIPKASSVDTELSEISTLDH	KIPKASSVRTELSDISTLDE
KIPKATSIPTELSAITTLYL	KIVKASSVPTFLSAISTLYL	KIPKASSVDTELSEISTLDR	KIPKASSVRTELSDISTLDH
KIPKATSFPTELSAISTLYL	KIVKASSVPTLLSAISTLYL	KIPKASSVDTELSEISTLED	KIPKASSVRTELSDISTLDR
KIPKATSPPTELSAISTLYL	KIVKASSVPTTLSAISTLYL	KIPKASSVDTELSEISTLEE	KIPKASSVRTELSDISTLED
KIPKATSTPTELSAISTLYL	KIVKASSVPTNLSAISTLYL	KIPKASSVDTELSEISTLEH	KIPKASSVRTELSDISTLEE
KIPKATSSPTELSAISTLYL	KIVKASSVPTVLSAISTLYL	KIPKASSVDTELSEISTLER	KIPKASSVRTELSDISTLEH
KIPKATSYPTELSAISTLYL	KIVKASSVPTILSAISTLYL	KIPKASSVDTELSEISTLHD	KIPKASSVRTELSDISTLER
KIPKATSIPTELSAISTLYL	KIIKASSVPTFLSAISTLYL	KIPKASSVDTELSEISTLHE	KIPKASSVRTELSDISTLHD

KIPKATSFPTELSAIYTLYL	KIIKASSVPTLLSAISTLYL	KIPKASSVDTELSEISTLHH	KIPKASSVRTELSDISTLHE
KIPKATSPPTELSAIYTLYL	KIIKASSVPTTLSAISTLYL	KIPKASSVDTELSEISTLHR	KIPKASSVRTELSDISTLHH
KIPKATSTPTELSAIYTLYL	KIIKASSVPTNLSAISTLYL	KIPKASSVDTELSEISTLRD	KIPKASSVRTELSDISTLHR
KIPKATSSPTELSAIYTLYL	KIIKASSVPTVLSAISTLYL	KIPKASSVDTELSEISTLRE	KIPKASSVRTELSDISTLRD
KIPKATSYPTELSAIYTLYL	KIIKASSVPTILSAISTLYL	KIPKASSVDTELSEISTLRH	KIPKASSVRTELSDISTLRE
KIPKATSIPTELSAIYTLYL	KIFKASSVPTFTSAISTLYL	KIPKASSVDTELSEISTLRR	KIPKASSVRTELSDISTLRH
KIPKATSFPTELSAIITLYL	KIFKASSVPTLTSAISTLYL	KIPKASSVDTELSHISTLDD	KIPKASSVRTELSDISTLRR
KIPKATSPPTELSAIITLYL	KIFKASSVPTTTSAISTLYL	KIPKASSVDTELSHISTLDE	KIPKASSVRTELSEISTLDD
KIPKATSTPTELSAIITLYL	KIFKASSVPTNTSAISTLYL	KIPKASSVDTELSHISTLDH	KIPKASSVRTELSEISTLDE
KIPKATSSPTELSAIITLYL	KIFKASSVPTVTSAISTLYL	KIPKASSVDTELSHISTLDR	KIPKASSVRTELSEISTLDH
KIPKATSYPTELSAIITLYL	KIFKASSVPTITSAISTLYL	KIPKASSVDTELSHISTLED	KIPKASSVRTELSEISTLDR
KIPKATSIPTELSAIITLYL	KILKASSVPTFTSAISTLYL	KIPKASSVDTELSHISTLEE	KIPKASSVRTELSEISTLED
KIPKASSPPTELSAIFTLYL	KILKASSVPTLTSAISTLYL	KIPKASSVDTELSHISTLEH	KIPKASSVRTELSEISTLEE
KIPKASSTPTELSAIFTLYL	KILKASSVPTTTSAISTLYL	KIPKASSVDTELSHISTLER	KIPKASSVRTELSEISTLEH
KIPKASSSPTELSAIFTLYL	KILKASSVPTNTSAISTLYL	KIPKASSVDTELSHISTLHD	KIPKASSVRTELSEISTLER
KIPKASSYPTELSAIFTLYL	KILKASSVPTVTSAISTLYL	KIPKASSVDTELSHISTLHE	KIPKASSVRTELSEISTLHD
KIPKASSIPTELSAIFTLYL	KILKASSVPTITSAISTLYL	KIPKASSVDTELSHISTLHH	KIPKASSVRTELSEISTLHE
KIPKASSFPTELSAIPTLYL	KITKASSVPTFTSAISTLYL	KIPKASSVDTELSHISTLHR	KIPKASSVRTELSEISTLHH
KIPKASSPPTELSAIPTLYL	KITKASSVPTLTSAISTLYL	KIPKASSVDTELSHISTLRD	KIPKASSVRTELSEISTLHR
KIPKASSTPTELSAIPTLYL	KITKASSVPTTTSAISTLYL	KIPKASSVDTELSHISTLRE	KIPKASSVRTELSEISTLRD
KIPKASSSPTELSAIPTLYL	KITKASSVPTNTSAISTLYL	KIPKASSVDTELSHISTLRH	KIPKASSVRTELSEISTLRE
KIPKASSYPTELSAIPTLYL	KITKASSVPTVTSAISTLYL	KIPKASSVDTELSHISTLRR	KIPKASSVRTELSEISTLRH
KIPKASSIPTELSAIPTLYL	KITKASSVPTITSAISTLYL	KIPKASSVDTELSRISTLDD	KIPKASSVRTELSEISTLRR
KIPKASSFPTELSAITTLYL	KINKASSVPTFTSAISTLYL	KIPKASSVDTELSRISTLDE	KIPKASSVRTELSHISTLDD
KIPKASSPPTELSAITTLYL	KINKASSVPTLTSAISTLYL	KIPKASSVDTELSRISTLDH	KIPKASSVRTELSHISTLDE
KIPKASSTPTELSAITTLYL	KINKASSVPTTTSAISTLYL	KIPKASSVDTELSRISTLDR	KIPKASSVRTELSHISTLDH
KIPKASSSPTELSAITTLYL	KINKASSVPTNTSAISTLYL	KIPKASSVDTELSRISTLED	KIPKASSVRTELSHISTLDR
KIPKASSYPTELSAITTLYL	KINKASSVPTVTSAISTLYL	KIPKASSVDTELSRISTLEE	KIPKASSVRTELSHISTLED
KIPKASSIPTELSAITTLYL	KINKASSVPTITSAISTLYL	KIPKASSVDTELSRISTLEH	KIPKASSVRTELSHISTLEE
KIPKASSFPTELSAIYTLYL	KIVKASSVPTFTSAISTLYL	KIPKASSVDTELSRISTLER	KIPKASSVRTELSHISTLEH
KIPKASSPPTELSAIYTLYL	KIVKASSVPTLTSAISTLYL	KIPKASSVDTELSRISTLHD	KIPKASSVRTELSHISTLER
KIPKASSTPTELSAIYTLYL	KIVKASSVPTTTSAISTLYL	KIPKASSVDTELSRISTLHE	KIPKASSVRTELSHISTLHD
KIPKASSSPTELSAIYTLYL	KIVKASSVPTNTSAISTLYL	KIPKASSVDTELSRISTLHH	KIPKASSVRTELSHISTLHE
KIPKASSYPTELSAIYTLYL	KIVKASSVPTVTSAISTLYL	KIPKASSVDTELSRISTLHR	KIPKASSVRTELSHISTLHH
KIPKASSIPTELSAIYTLYL	KIVKASSVPTITSAISTLYL	KIPKASSVDTELSRISTLRD	KIPKASSVRTELSHISTLHR
KIPKAYSFPTELSAIFTLYL	KIIKASSVPTFTSAISTLYL	KIPKASSVDTELSRISTLRE	KIPKASSVRTELSHISTLRD
KIPKAYSPPTELSAIFTLYL	KIIKASSVPTLTSAISTLYL	KIPKASSVDTELSRISTLRH	KIPKASSVRTELSHISTLRE
KIPKAYSTPTELSAIFTLYL	KIIKASSVPTTTSAISTLYL	KIPKASSVDTELSRISTLRR	KIPKASSVRTELSHISTLRH
KIPKAYSSPTELSAIFTLYL	KIIKASSVPTNTSAISTLYL	KIPKASSVETELSDISTLDD	KIPKASSVRTELSHISTLRR
KIPKAYSYPTELSAIFTLYL	KIIKASSVPTVTSAISTLYL	KIPKASSVETELSDISTLDE	KIPKASSVRTELSRISTLDD
KIPKAYSIPTELSAIFTLYL	KIIKASSVPTITSAISTLYL	KIPKASSVETELSDISTLDH	KIPKASSVRTELSRISTLDE
KIPKAYSFPTELSAIPTLYL	KIFKASSVPTFNSAISTLYL	KIPKASSVETELSDISTLDR	KIPKASSVRTELSRISTLDH
KIPKAYSPPTELSAIPTLYL	KIFKASSVPTLNSAISTLYL	KIPKASSVETELSDISTLED	KIPKASSVRTELSRISTLDR
KIPKAYSTPTELSAIPTLYL	KIFKASSVPTTNSAISTLYL	KIPKASSVETELSDISTLEE	KIPKASSVRTELSRISTLED
KIPKAYSSPTELSAIPTLYL	KIFKASSVPTNNSAISTLYL	KIPKASSVETELSDISTLEH	KIPKASSVRTELSRISTLEE
KIPKAYSYPTELSAIPTLYL	KIFKASSVPTVNSAISTLYL	KIPKASSVETELSDISTLER	KIPKASSVRTELSRISTLEH
KIPKAYSIPTELSAIPTLYL	KIFKASSVPTINSAISTLYL	KIPKASSVETELSDISTLHD	KIPKASSVRTELSRISTLER
KIPKAYSFPTELSAITTLYL	KILKASSVPTFNSAISTLYL	KIPKASSVETELSDISTLHE	KIPKASSVRTELSRISTLHD
KIPKAYSPPTELSAITTLYL	KILKASSVPTLNSAISTLYL	KIPKASSVETELSDISTLHH	KIPKASSVRTELSRISTLHE
KIPKAYSTPTELSAITTLYL	KILKASSVPTTNSAISTLYL	KIPKASSVETELSDISTLHR	KIPKASSVRTELSRISTLHH
KIPKAYSSPTELSAITTLYL	KILKASSVPTNNSAISTLYL	KIPKASSVETELSDISTLRD	KIPKASSVRTELSRISTLHR
KIPKAYSYPTELSAITTLYL	KILKASSVPTVNSAISTLYL	KIPKASSVETELSDISTLRE	KIPKASSVRTELSRISTLRD
KIPKAYSIPTELSAITTLYL	KILKASSVPTINSAISTLYL	KIPKASSVETELSDISTLRH	KIPKASSVRTELSRISTLRE
KIPKAYSFPTELSAISTLYL	KITKASSVPTFNSAISTLYL	KIPKASSVETELSDISTLRR	KIPKASSVRTELSRISTLRH
KIPKAYSPPTELSAISTLYL	KITKASSVPTLNSAISTLYL	KIPKASSVETELSEISTLDD	
KIPKAYSTPTELSAISTLYL	KITKASSVPTTNSAISTLYL	KIPKASSVETELSEISTLDE	
KIPKAYSSPTELSAISTLYL	KITKASSVPTNNSAISTLYL	KIPKASSVETELSEISTLDH	

KIPKAYSYPTELSAISTLYL	KITKASSVPTVNSAISTLYL	KIPKASSVETELSEISTLDR	
KIPKAYSIPTELSAISTLYL	KITKASSVPTINSAISTLYL	KIPKASSVETELSEISTLED	
KIPKAYSFPTELSAIYTLYL	KINKASSVPTFNSAISTLYL	KIPKASSVETELSEISTLEE	
KIPKAYSPPTELSAIYTLYL	KINKASSVPTLNSAISTLYL	KIPKASSVETELSEISTLEH	
KIPKAYSTPTELSAIYTLYL	KINKASSVPTTNSAISTLYL	KIPKASSVETELSEISTLER	
KIPKAYSSPTELSAIYTLYL	KINKASSVPTNNSAISTLYL	KIPKASSVETELSEISTLHD	
KIPKAYSYPTELSAIYTLYL	KINKASSVPTVNSAISTLYL	KIPKASSVETELSEISTLHE	

SUPPLEMENTARY TABLE 2: Search range of hyperparameters for different ML models using grid search.

ML Model	Hyperparameters	Search Range
BL	NA	Not applicable
RR	alpha	0-1 with increments of 0.05
LR	alpha	0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000
	max_iter	range(1,1000)
EN	alpha	0.001, 0.01, 0.1, 1
	max_iter	Range (1,1000)
	tol	0.0001, 0.001, 0.01
	l1_ratio	0.05-0.95 with increments of 0.05
SVR	kernel	'linear', 'poly', 'rbf'
	gamma	'scale'
	C	1, 2, 3, 4, 5
KRR	alpha	0.001, 0.01, 0.1, 1, 10
	kernel	'poly', 'rbf', 'sigmoid'
	gamma	0.001, 0.01, 0.1, 1, 10
	coef0	0, 0.1, 0.5, 1
	degree	2, 3, 4, 5
RF	max_depth	None, 5, 10, 15, 20
	min_samples_leaf	2, 4, 8, 16, 32
	max_features	'sqrt', 'log2'
	n_estimators	100, 200, 300, 400
	bootstrap	True, False
	min_samples_split	5, 10, 15, 20, 25
NN	hidden_layer_sizes	(10)-(200), (10,10)-(200,200), (10,10,10)-(200,200,200) with increments of 10, 50, 100, and 200
	max_iter	100, 200, 500, 1000, 2000, 5000, 10000
	solver	'adam', 'sgd', 'lbfgs'
	activation	'relu', 'tanh', 'logistic'
	learning_rate	'adaptive', 'constant'
	alpha	0.001, 0.01, 0.1, 1, 10
	learning_rate_init	0.0001, 0.001, 0.01, 0.1

SUPPLEMENTARY TABLE 3: List of peptide sequences used to determine R^2 scores of the R_g of sequences in PDB databank to determine the natural variance of the R_g values.

Peptide sequence	# of residues	PDB ID	R_g of structure with least energy(in Å)
SWTWENGKWTWK	12	1LE1	5.84
GSKYSDTADESSYRW	15	7LCW	5.15
GKVLSKIFGN	10	2bau	4.14
GCCSDPRCAWRC	12	1im1	5.04
MENTSITIEFSSKF	14	2NB7	6.32
KKLLPIVANLLKSLL	15	6gij	7.48
DPRIAARGRARLRAVGAI	18	2lx5	8.74
YSDELRQRLAARLEALKENG	20	1odr	9.56
KMNMLKENVDYIQKNQNLFK	20	2mu9	8.84
SMHPGHLKGREIGMWYAKKQ	20	2n16	7.76
SDTRYNKDFINNKHLNEHAH	20	5uy2	7.3
RWKRHISEQLRRRDRLQRQA	20	5zyx	7.51
SSMKLSFRARAYGFRGPGPQL	21	1lv4	6.33
SRPLIHFGNDYEDRYYRENMY	21	1s4t	7.7
FFHHIFRGIVHVGKTIHRLVTG	21	2ojm	9.73
PLVEDMQRQWAGLVEKVQAAVGT	23	1oeg	10.22
SWFEPLVEDMQRQWAGLVEKVQAA	24	1oef	10.85
MPGTIKENIIGVSYDEYRYRSVIKA	25	1ckw	11.1
HCDAATICPDGTTCSLSPYGVWYCSPFS	28	1i8y	8.86
MKTILRFVAGYDIASHKKKTGGYPWERGKA	30	2mlu	10.5
MAMAMRSTFAARVGAKPAVRGARPASRMSCMA	32	1fct	10.68
ESRKGQERFNRWFLTGMTVAGVVLLGSLFSRK	32	6f46	13.79
YKFACPECPKRFMRSDHLSKHITLHELLGEERR	33	1RIM	8.52
KKWGWLAWVDPAYEFIKGFGKGAIKEGNKDKWKNI	35	2jpk	11.79
FTLSLDVPTNIMNLLFNIAKAKNLRAQAAANAHLMAQI	38	2rmh	13.84
TVFAFASMLCLLNSTVNPIIYALRSKDLRHAFRSMFPSAE	40	2koe	11.28
ATCDLASGFGVGSSLCAAHCLVKGYRGGYCKNKICHCRDKF	41	2ny9	9.38
GKIPVKAIKQAGKVIGKGLRAINIAGTTHDVVSFFRPKKKKH	42	2jr8	18.53
MTISQQEFGRTGLPDLSSMTEEEQIAYAMQMSLQGAEFGQAESAD	45	1p9c	15.12
MEALSWDCFALSLLAVALGLGVPMLHHLCGWDLWYCFHLCLAWLPWRGRQ	50	8ar3	15.56
GPLGSIKELKMSKDEIKREYKEMEGSPEIKSKRRQFHQEIQSRNMRENVKRSSVVVAN	58	2ml9	22.56
SSGHIEGRHMNKQATEILNEYFYSHLSNPYPSEEAKEELAKKCGITVSQVSNWFGNKRIRYKKN	64	1du6	11.66
TFLTQVKESLSSYWESAKTAAQNLYEKTYLPAVDEKLRDLYSKSTAAMSTYTGIFTDQVLSVLKGEE	67	1i5j	16.04
MRPEVASTFKVLRNVTVVLWSAYPVVVWLIGSEGAGIVPLNIETLLFMVLDVSAKVGFGLILLRSRAIFG	69	1bct	28.28

SUPPLEMENTARY TABLE 4: Tukey's post hoc test results of R^2 scores of the models predicting $\langle R_g \rangle$ using different feature scales.

Group	Control Group	Lower Limit	Difference	Upper Limit
z-scale	t-scale	-0.033761358	-0.00125	0.031261358
z-scale	nominal scale	0.057488642	0.09	0.122511358
t-scale	nominal scale	0.058738642	0.09125	0.123761358

SUPPLEMENTARY TABLE 5: Tukey's post hoc test results of R2 scores of the models predicting <EtE> using different feature scales.

Group	Control Group	Lower Limit	Difference	Upper Limit
z-scale	t-scale	-0.031986947	0.005	0.041986947
z-scale	nominal scale	0.046763053	0.08375	0.120736947
t-scale	nominal scale	0.041763053	0.07875	0.115736947

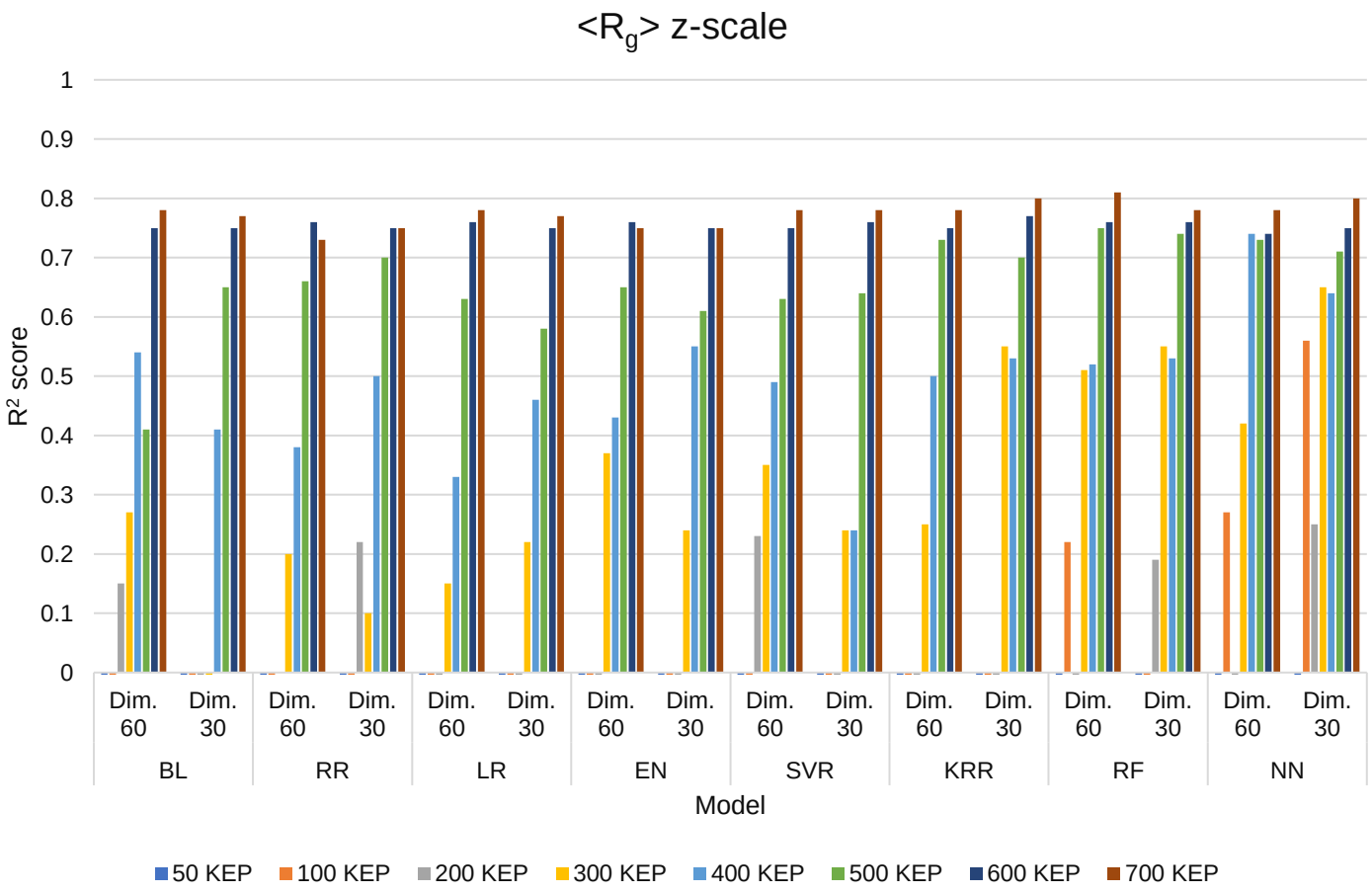
SUPPLEMENTARY TABLE 6: Mean Square error (MSE), Pearsons's Correlation Coefficient (PCC) and Mean Absolute Error (MAE) of the best performing models predicting $<R_g>$.

Model	MSE	PCC	MAE
z-scale; RF	0.601	0.903	0.467
t-scale; RF	0.610	0.901	0.475
reduced z-scale; RF	0.611	0.902	0.466
reduced t-scale; RF	0.615	0.903	0.472
normalized z-scale; RF	0.607	0.900	0.474
normalized t-scale; RF	0.618	0.899	0.475

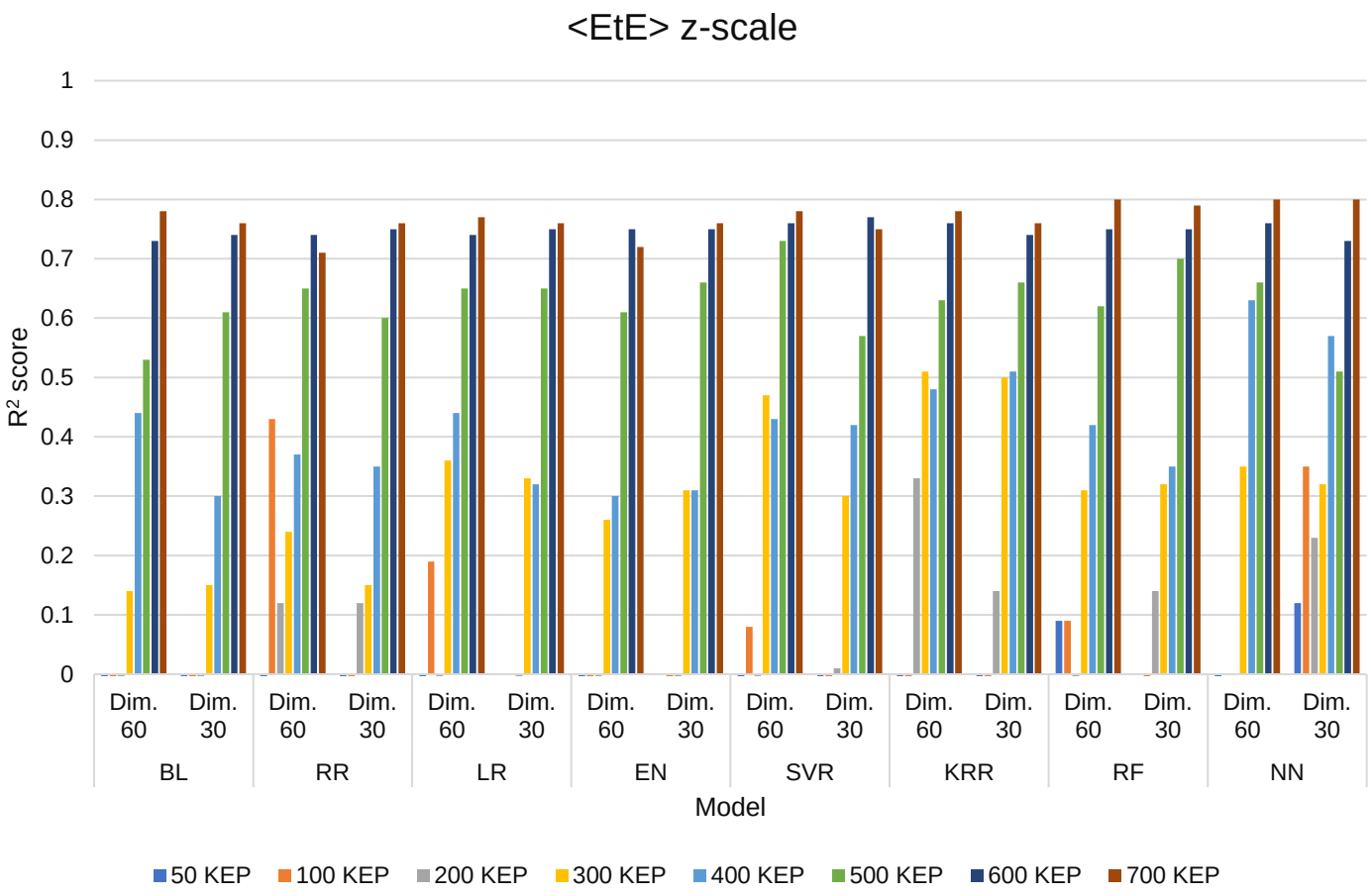
SUPPLEMENTARY TABLE 7: Mean Square error (MSE), Pearsons's Correlation Coefficient (PCC) and Mean Absolute Error (MAE) of the best performing models predicting <EtE>.

Model	MSE	PCC	MAE
z-scale; RF	2.72	0.897	2.13
t-scale; RF	2.72	0.892	2.16
normalized z-scale; RF	2.71	0.898	2.13

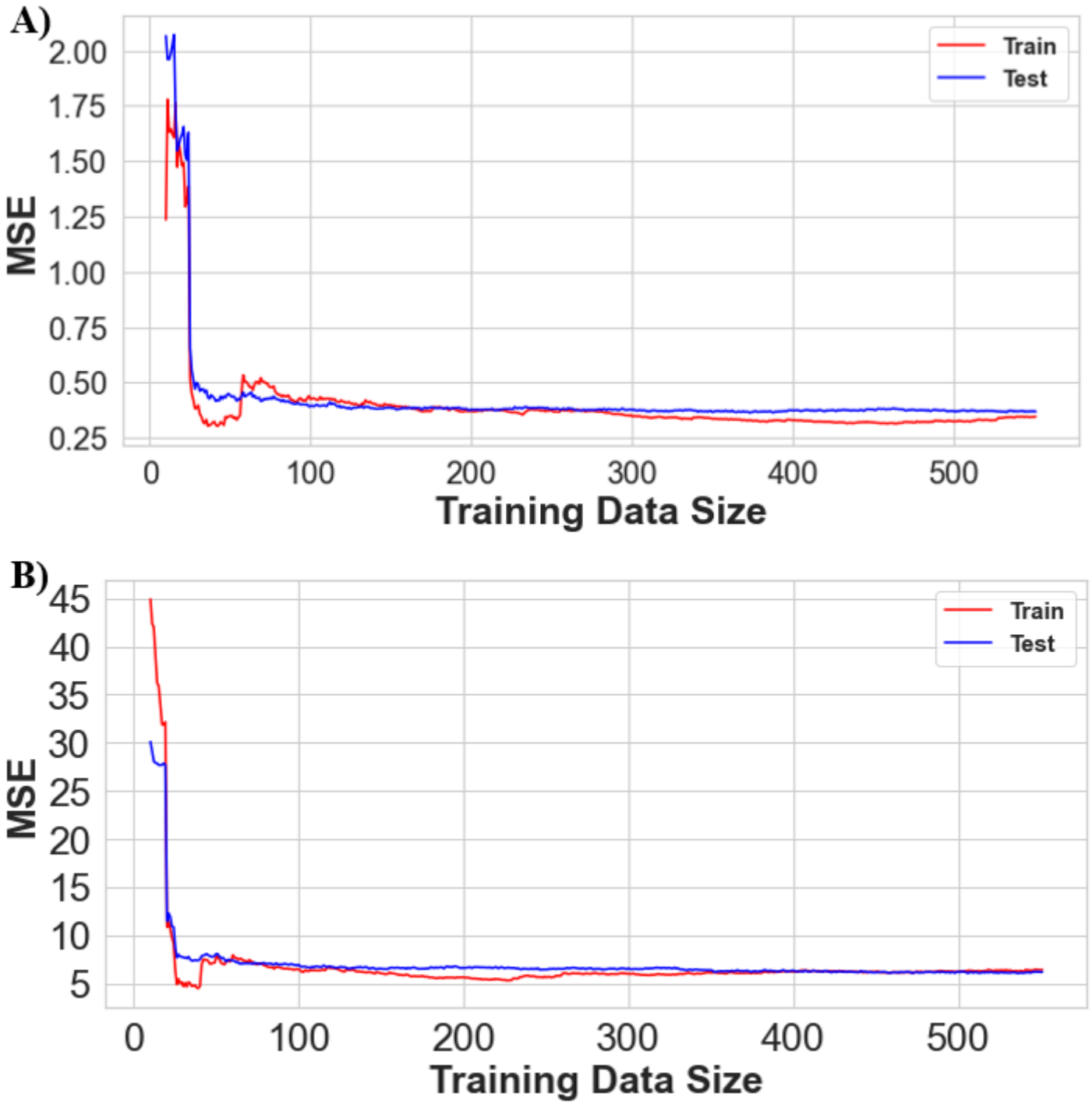
SUPPLEMENTARY FIGURE 1: R^2 scores of the models predicting <Rg> using z-scale to describe the residues of the sequences as a function of increasing the number of datapoints. 'Dim 60' stands for no dimensionality reduction using the z-scale and 'Dim 30' refers to dimensionality reduction by half from 60 to 30.



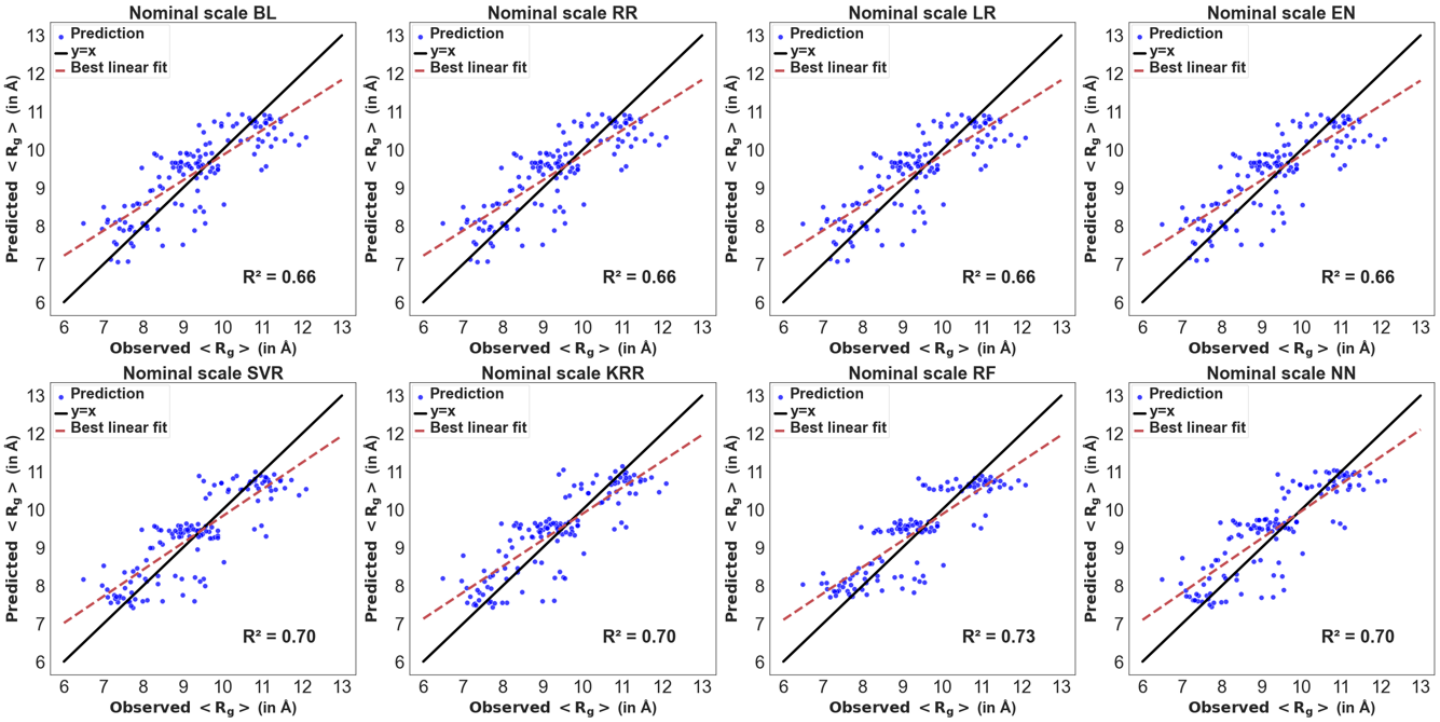
SUPPLEMENTARY FIGURE 2: R^2 scores of the models predicting <EtE> using z-scale to describe the residues of the sequences as a function of increasing the number of datapoints. 'Dim 60' stands for no dimensionality reduction using the z-scale and 'Dim 30' refers to dimensionality reduction by half from 60 to 30.



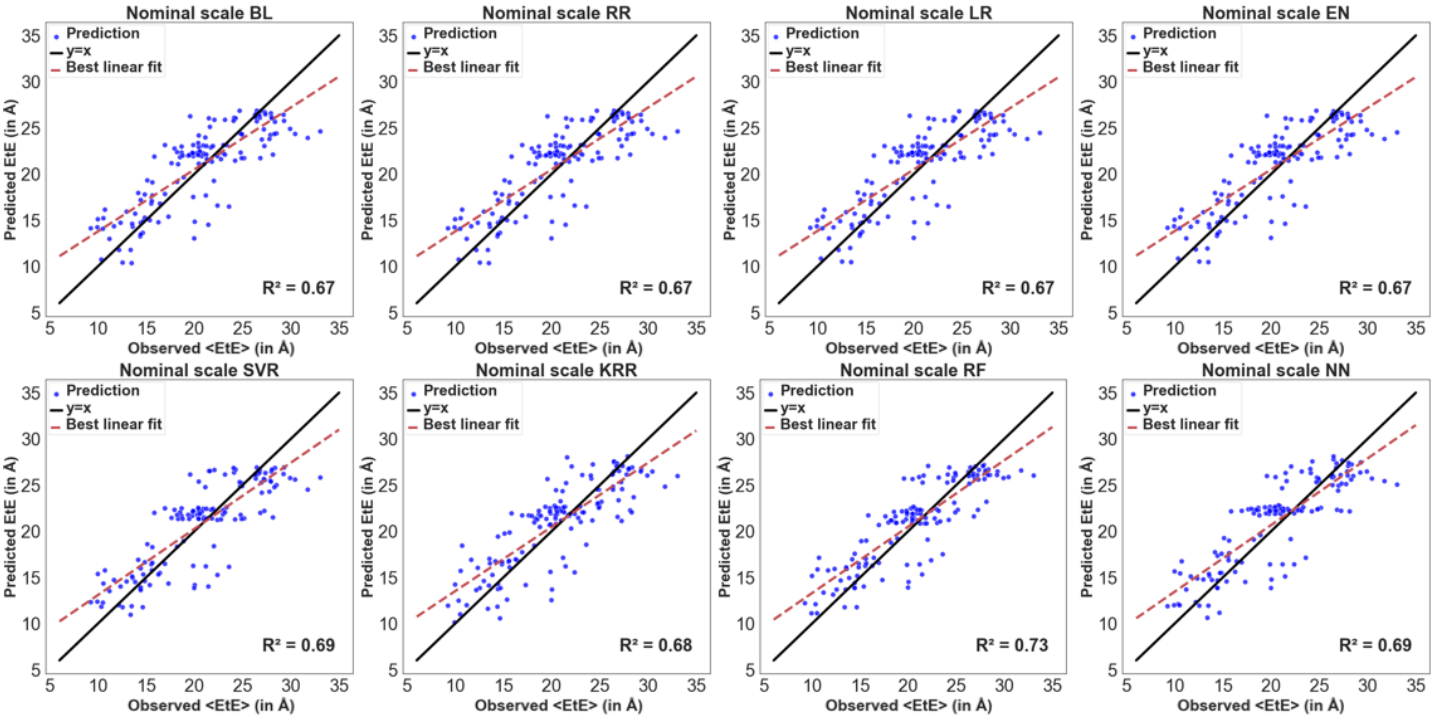
SUPPLEMENTARY FIGURE 3: Learning curves of the best performing models. A) Learning curve of RF model with z-scale predicting <R_g>. B) Learning curve of RF model with normalized z-scale predicting <EtE>.



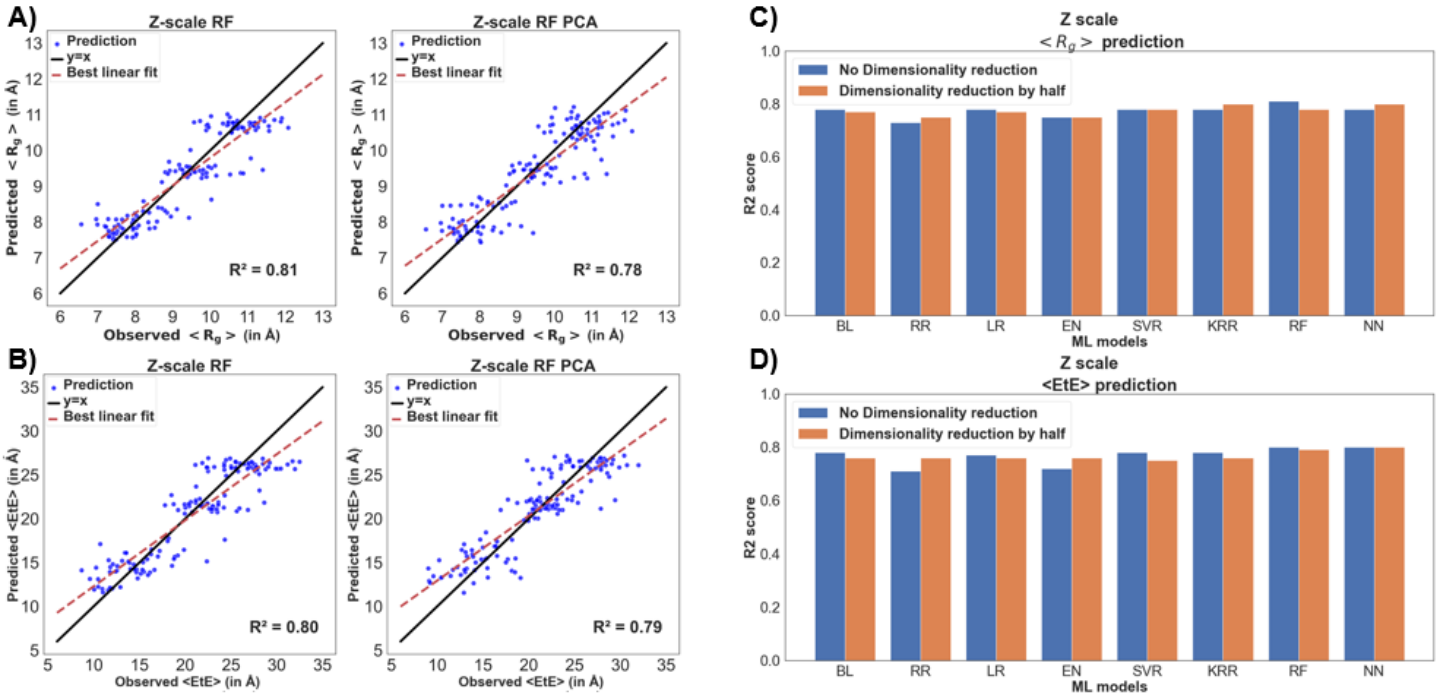
SUPPLEMENTARY FIGURE 4: **Parity plots of all models predicting <R_g> built with nominal scale describing the residues in the sequences.**



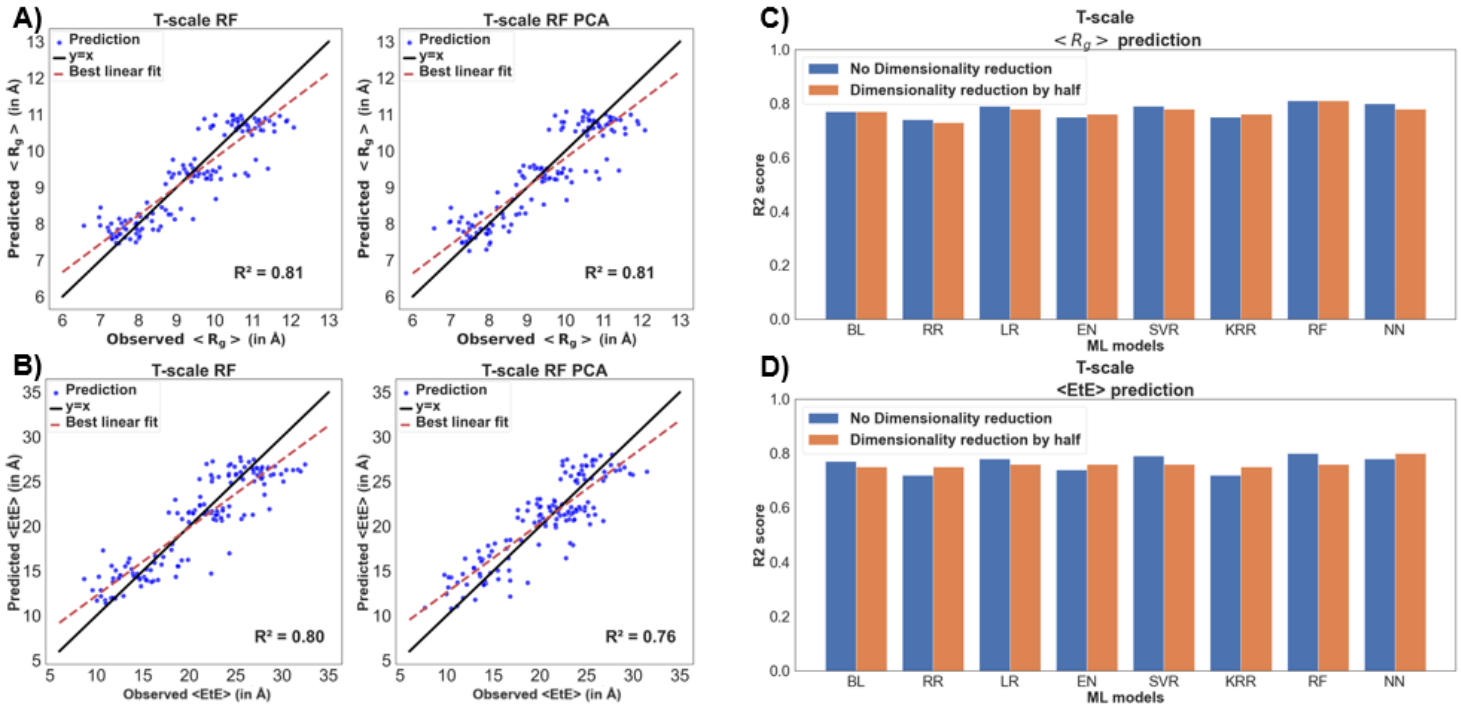
SUPPLEMENTARY FIGURE 5: Parity plots of all models predicting <EtE> built with nominal describing the residues in the sequences.



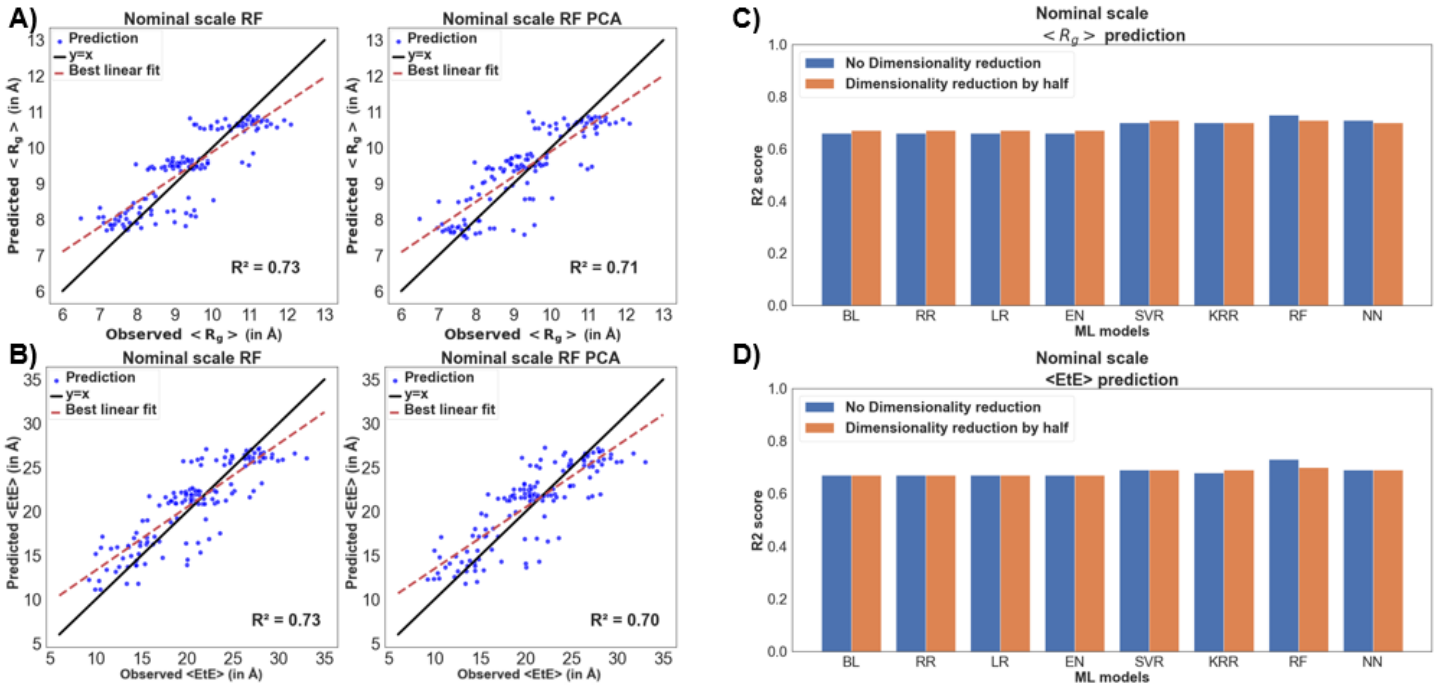
SUPPLEMENTARY FIGURE 6: Comparison of the performance of the models before and after dimensionality reduction built with z-scale features.



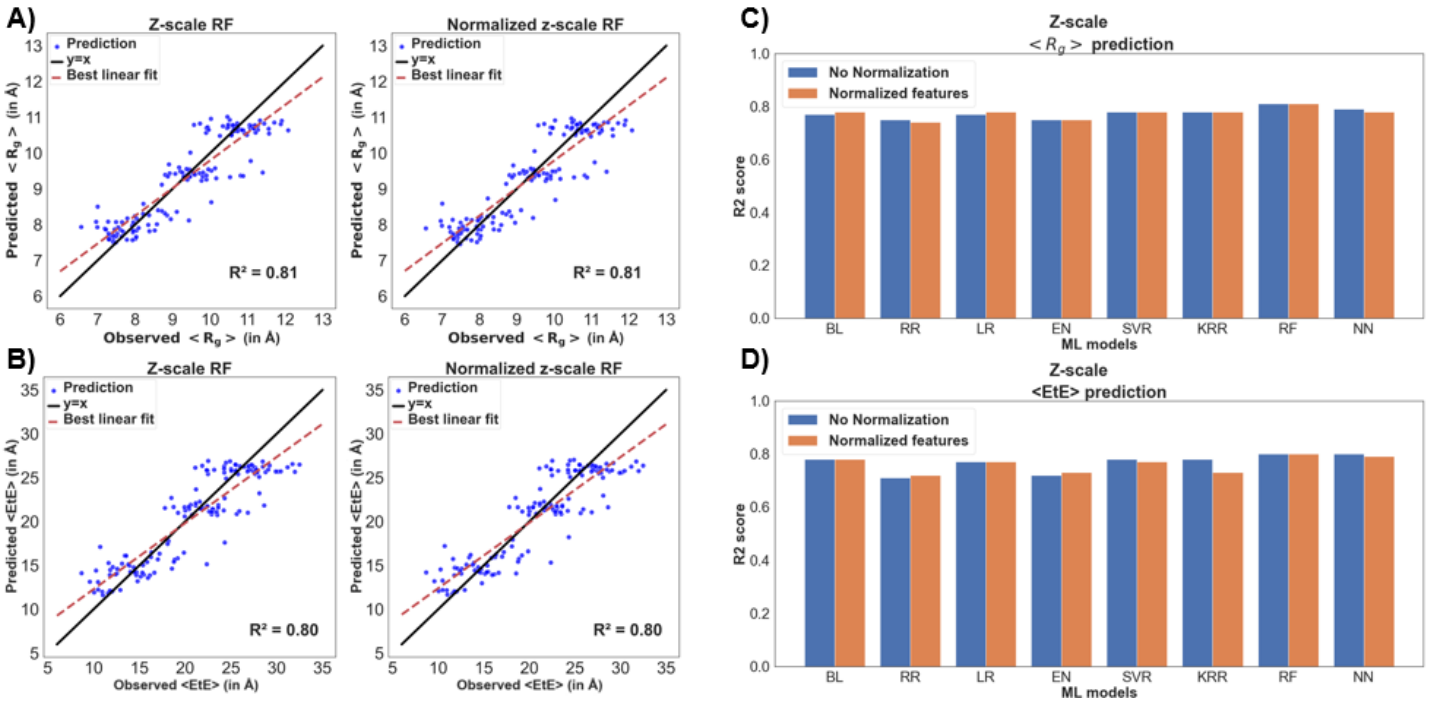
SUPPLEMENTARY FIGURE 7: Comparison of the performance of the models before and after dimensionality reduction built with t-scale features.



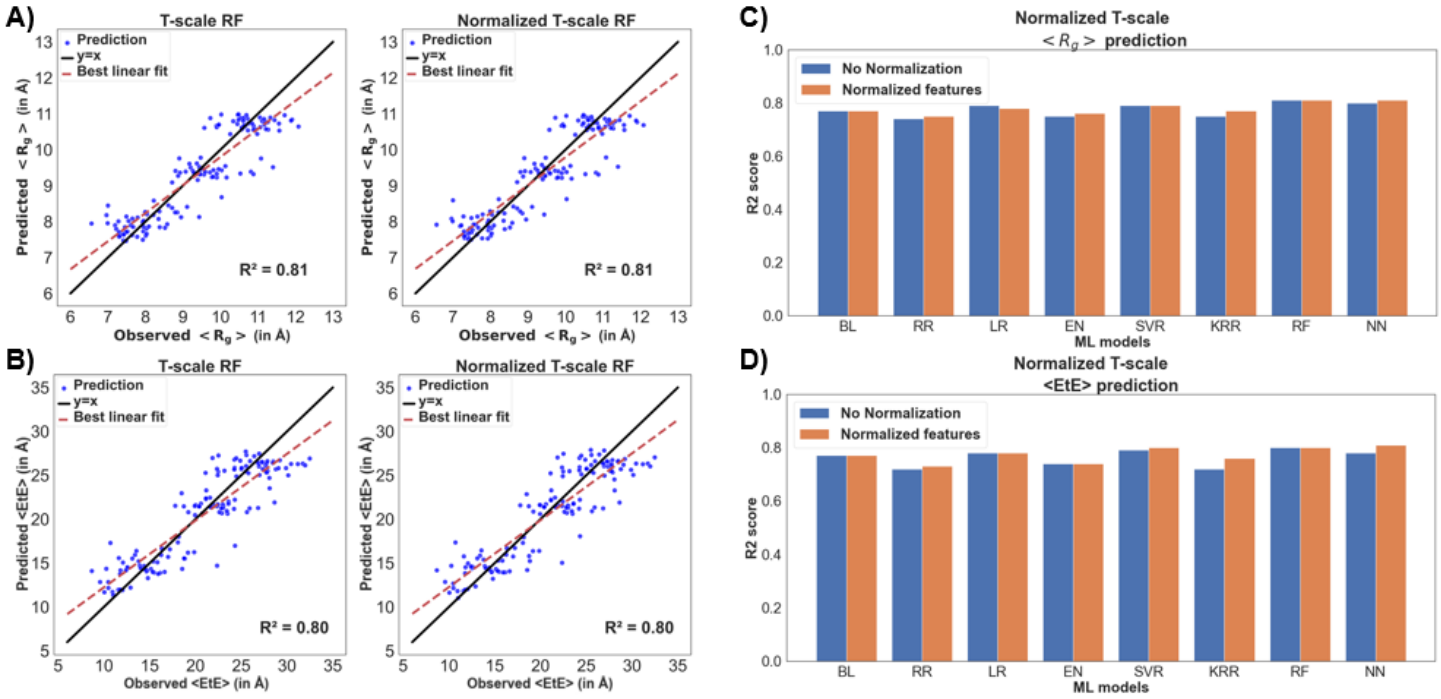
SUPPLEMENTARY FIGURE 8: Comparison of the performance of the models before and after dimensionality reduction built with nominal scale features.



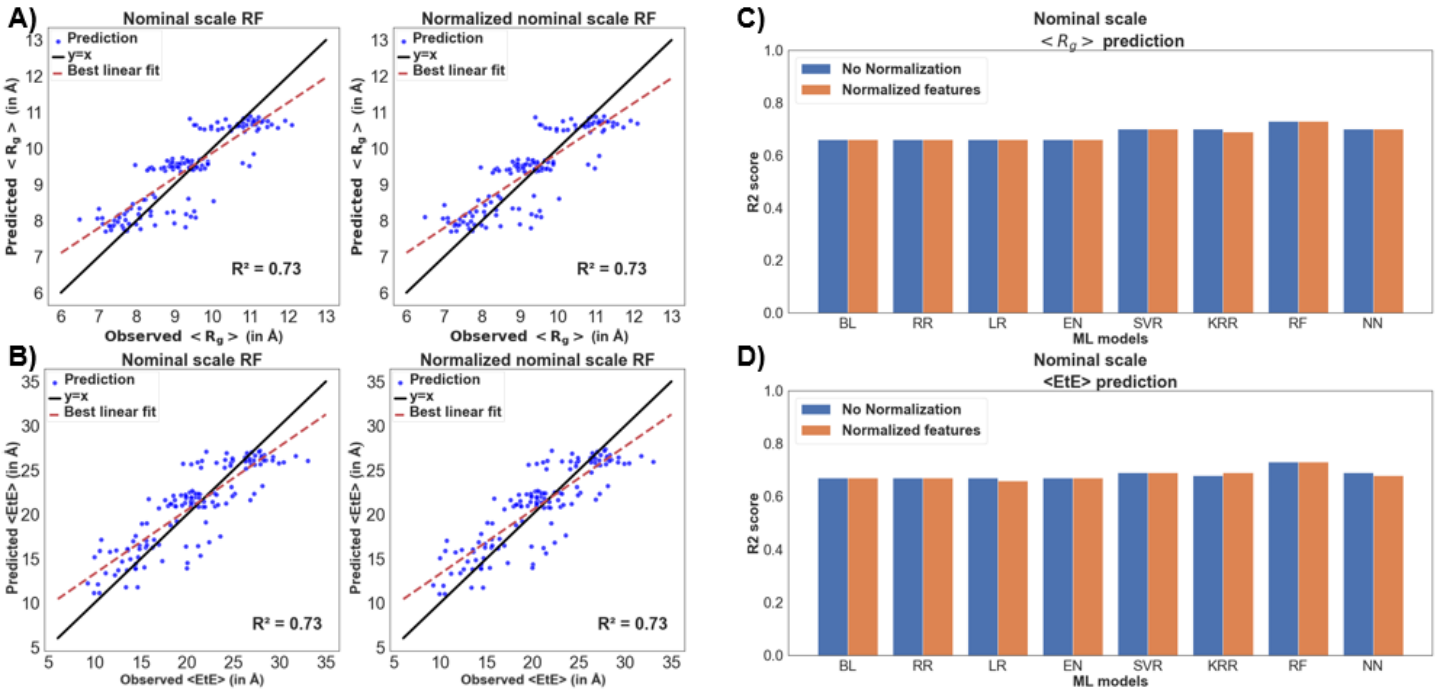
SUPPLEMENTARY FIGURE 9: Comparison of the performance of the models before and after normalization built with z-scale features.



SUPPLEMENTARY FIGURE 10: **Comparison of the performance of the models before and after normalization built with t-scale features.**



SUPPLEMENTARY FIGURE 11: Comparison of the performance of the models before and after normalization built with nominal scale features.



SUPPLEMENTARY CODE 1: Perl script to perform geometry optimization of the mesostructure

```
#!perl

#---------------------------------------
#Geometry optimization task Perl script
#---------------------------------------

use strict;
use Getopt::Long;
use MaterialsScript qw(:all);

my $doc = $Documents{"mesostructure.xsd"}; # Enter the name of your mesostructure file  here (in.xsd format)

my $results = Modules->Mesocite->GeometryOptimization->Run($doc, Settings(
    '3DPeriodicvdWSummationMethod' => 'Ewald',
    '3DPeriodicElectrostaticSummationMethod' => 'PPPM',
    '3DPeriodicvdWEwaldRepulsiveCutOff' => 6.46,
    CurrentForcefield => '/Forcefield', # enter the name of the forcefield (file type will be in .off format)
    MaxIterations => 1000,
    UseMaxDisplacement => 'Yes',
    UseMaxEnergy => 'No'));
```

SUPPLEMENTARY CODE 2: Perl script to perform Dissipative Particle Dynamics of the mesostructure

```
#!perl

#-----------------------------------------------
#Dissipative particle Dynamics task Perl script
#-----------------------------------------------

use strict;
use Getopt::Long;
use MaterialsScript qw(:all);

my $doc = $Documents{"mesostructure.xsd"}; # Enter the name of your mesostructure file  here (in.xsd format)

my $results = Modules->Mesocite->DPD->Run($doc, Settings(
    '3DPeriodicvdWSummationMethod' => 'Ewald',
    '3DPeriodicElectrostaticSummationMethod' => 'PPPM',
    '3DPeriodicvdWEwaldRepulsiveCutOff' => 6.46,
    CurrentForcefield => '/Forcefield', # enter the name of the forcefield (file type will be in .off format)
    NumberOfSteps => 666000,
    TimeStep => 15.079,
    TrajectoryFrequency => 33300,
    Dissipation => 0.0805756,
    DissipationRadius => 6.46));
my $outTrajectory = $results->Trajectory;
```