



T.C.

GAZİANTEP ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**PROTEİN KATLANMASI DİNAMİĞİ:
TRANSFER EDİLEBİLİR ETKİLEŞİM POTANSİYELİ DİZAYNI ve
FARKLI KOMPLEKSLİK SEVİYELERİNE SAHİP YENİ MONTE
CARLO ve MOLEKÜLER DİNAMİK PROTEİN MODELLERİNİN
GELİŞTİRİLMESİ**

Gökhan SELAMET

DOKTORA TEZİ

BİYOENFORMATİK ve BİLİŞİMSEL BİYOLOJİ ANABİLİM DALI
BİYOENFORMATİK ve BİLİŞİMSEL BİYOLOJİ DOKTORA PROGRAMI

TEZ DANIŞMANI

Doç. Dr. Tuğba TAŞKIN TOK

II. TEZ DANIŞMANI

Prof. Dr. Hüseyin KAYA

Gaziantep

2023

T.C.
GAZİANTEP ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOENFORMATİK ve BİLİŞİMSEL BİYOLOJİ ANABİLİM DALI
Biyoenformatik ve Bilişimsel Biyoloji Doktora Programı

PROTEİN KATLANMASI DİNAMİĞİ:
TRANSFER EDİLEBİLİR ETKİLEŞİM POTANSİYELİ DİZAYNI ve FARKLI
KOMPLEKSLİK SEVİYELERİNE SAHİP YENİ MONTE CARLO ve MOLEKÜLER
DİNAMİK PROTEİN MODELLERİNİN GELİŞTİRİLMESİ

Gökhan SELAMET

Tez Savunma Tarihi : 21/06/2023

Doç. Dr. Davut Sinan KAPLAN

Sağlık Bilimleri Enstitüsü Müdürü

Sağlık Bilimleri Enstitüsü Müdürlüğü'ne sunulan bu çalışmanın; bir "Doktora" derecesi için uygun ve yeterli bir çalışma olduğunu onaylıyorum.

Doç. Dr. Tuğba TAŞKIN TOK

Biyoenformatik ve Bilişimsel Biyoloji Anabilim Dalı Başkanı

Bu tez tarafımda okunmuş, kapsamı ve niteliği açısından bir "Doktora" tezi olarak kabul edilmiştir.

Doç. Dr. Tuğba TAŞKIN TOK
Tez Danışmanı

Prof. Dr. Hüseyin KAYA
II. Tez Danışmanı

Bu tez tarafımda okunmuş, kapsamı ve niteliği açısından bir "Doktora" tezi olarak kabul edilmiştir.

Tez Jürisi

İmza

Prof. Dr. Ahmet BİNGÜL

.....

Prof. Dr. Murat CANPOLAT

.....

Prof. Dr. Zafer ÇETİN

.....

Doç. Dr. Ömer Faruk CİHAN

.....

Doç. Dr. Tuğba TAŞKIN TOK

.....



T.C.

GAZİANTEP ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**PROTEİN KATLANMASI DİNAMİĞİ:
TRANSFER EDİLEBİLİR ETKİLEŞİM POTANSİYELİ DİZAYNI ve
FARKLI KOMPLEKS LİK SEVİYELERİNE SAHİP YENİ MONTE
CARLO ve MOLEKÜLER DİNAMİK PROTEİN MODELLERİNİN
GELİŞTİRİLMESİ**

Gökhan SELAMET

DOKTORA TEZİ

BİYOENFORMATİK ve BİLİŞİMSEL BİYOLOJİ ANABİLİM DALI
BİYOENFORMATİK ve BİLİŞİMSEL BİYOLOJİ DOKTORA PROGRAMI

TEZ DANIŞMANI

Doç. Dr. Tuğba TAŞKIN TOK

II. TEZ DANIŞMANI

Prof. Dr. Hüseyin KAYA

Gaziantep

2023

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığı beyan ederim.

Tarih: 21/06/2023

Gökhan SELAMET

TEŞEKKÜR

Bu doktora tez çalışmasında yer alan kısmi nümerik hesaplamalar TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirilmiştir. Ayrıca Gaziantep Üniversitesi, Biyoenformatik ve Bilişimsel Biyoloji araştırma laboratuvarının hesaplama kapasitesi

Tez çalışmasının oluşumunda, birçok kişiden büyük destek ve ilham aldım. Öncelikle, tez danışmanlarım Doç. Dr. Tuğba Taşkın Tok ve Prof Dr. Hüseyin Kaya tez sürecimde ki rehberlerim ve kılavuzumdur. Tez danışmanlarımla bilgeliği, yol gösterici eleştirileri ve cesaretlendirici sözleri, bu tezin olgunlaşmasında büyük bir rol oynadı.

Bu tez çalışması sırasında programlama üzerine yaptığımız tartışmalar ve fikir alışverişleri ile her zaman yardımlarını ve desteklerini esirgemeyen, değerli abim Ferhat Ayaz'a teşekkür bir borç bilirim.

Aileme, arkadaşlarıma ve sevdiklerime, bu tez çalışmasında verdikleri sonsuz destek ve anlayışları için minnettarım. Onların cesaretlendirici sözleri ve sabırları, benim bu karmaşık çalışma sürecinde devam etmemi sağladı.

Son olarak, hayatımda derin izler bırakan ve bilim sevgisini bana genç bir lise öğrenciyken aşılayan rahmetli Mustafa Sertbel öğretmenimi anmadan geçemiyeceğim. Onun bana kattığı değerli öğretileri ve yaşam boyu sürececek ilhamı hala benimledir. Bu tez çalışması, onun anısını onurlandırmak ve onunla geçirdiğim zamanın değerini göstermek için bir fırsattır. Bu tez çalışmasının tamamlanmasında ve bugünkü başarılarımda onun etkisi büyüktür. Minnettarlığımı ifade etmek için bu teşekkür paragrafını sunuyorum ve ruhunun huzur içinde olduğunu umuyorum.

İÇİNDEKİLER

	<u>Sayfa</u>
KISALTMA ve SİMGELER.....	vi
TABLO LİSTESİ	viii
ŞEKİL LİSTESİ.....	ix
ÖZET	xii
SUMMARY	xiii
1. GİRİŞ ve AMAÇ.....	1
2. GENEL BİLGİLER	3
2.1 Protein Katlanması	3
2.2 Protein Katlanma Dinamiğinin Modellenmesi ve Yapı Tahminlerinin Gelişimi	4
2.3 Protein Spesifik Model Yaklaşımları: Gō Modeli	4
2.3.1 Amino Asitlerin C_{α} Atomları ile Temsil Edildiği Kabalaştırılmış Protein Moleküler Dinamik Gō Modeli	6
2.3.2 Model Konformasyonel Enerji Bileşenleri.....	6
2.3.2.1 Yerel - Açısal etkileşimler.....	7
2.3.2.2 Gerilme terimi.....	8
2.3.2.3 Bükülme terimi	9
2.3.2.4 Burulma Enerji Terimi	9
2.3.3 Yerel Olmayan - Bağırsız Etkileşimler.....	10
2.3.3.1 Sterik etkileşim enerjisi.....	11
2.3.3.2 Lennard-Jones (LJ) etkileşim enerjisi	11
2.4 Proteinlerde termodinamik kooperativite	13
2.5 Protein katlanma kinetiği.....	13
3. GEREÇ ve YÖNTEM	15
3.1 Moleküler Dinamik Simülasyon Programı	15
3.1.1 Moleküler Dinamik Motoru: GoMDM	15
3.1.2 GoMDM'nin Kullanılması	16
3.1.3 GoMDM'in İşlevi	17
3.1.4 Parçacıklar arası mesafe hesabı	18
3.1.5 pdb Okuyucu: cpdb	19
3.1.6 cpdb Kullanımı	20
3.1.7 Kodların erişimi	20
3.2 Protein İçi Etkileşim Ağları.....	20
3.2.1 Protein içi etkileşim ağlarının oluşturulması.....	21
3.2.2 Moleküler dinamik simülasyonlarında kullanılan protein kümesi	22

4. BULGULAR.....	23
4.1 Protein İçi Etkileşim Ağlarının Protein Katlanmasına Etkileri	23
4.1.1 Protein İçi Etkileşim Ağlarının Protein Katlanma Termodinamiğine Etkisi	23
4.1.1.1 Etkileşim Sona Erim Mesafe Kriteri ve Termodinamik Kooperativite İlişkisi.....	25
4.1.1.2 Ortalama etkileşim sayısı ve termodinamik kooperativite ilişkisi	27
4.1.2 Etkileşim ağlarının protein katlanma kinetiğine etkisi	28
4.1.2.1 Literatürdeki etkileşim düzen parametreleri ve katlanma oranları ilişkisi.....	29
4.1.3 Açısal Etkileşimli Etkileşim Düzen Parametresi.....	32
4.1.3.1 AEEDP etkileşim düzen parameteresi ve katlanma oranları arasındaki ilişki.....	33
4.1.3.2 AEEDP etkileşim düzen parametresi ve termodinamik kooperativite ilişkisi.....	35
4.2 Tüm Atomların Temsil Edildiği Monte Carlo Protein Modeli	36
4.2.1 Transfer Edilebilir İstatistiksel Etkileşim Potansiyel Dizaynları	37
4.2.1.1 Atom çiftleri için transfer edilebilir istatistiksel etkileşim enerji terimi dizaynı	39
4.2.2 Yönelim ve düzlemsellik özelliği olan transfer edilebilir istatistiksel hidrojen bağı enerji terimi dizaynı	42
4.2.3 Protein zinciri üzerinde yerel olan amino asit üçlüleri ve dörtlüleri için transfer edilebilir istatistiksel etkileşim enerji terimleri dizaynı	47
4.2.3.1 $C_\alpha - C_\beta$ Kovalent Bağlarının Birbirine Göre Hizalanmasını η Dihedral ve θ Bağ Açılarını ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\eta-\theta}$	50
4.2.3.2 $C = O$ Kovalent Bağlarının Birbirine Göre Hizalanmasını ξ Dihedral ve $\tilde{\theta}$ Bağ Açılarını ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\xi-\tilde{\theta}}$	53
4.2.3.3 Amino asit üçlüsünde tanımlanan ψ ve ϕ Dihedral Açılarını ile Tasvir Edilen Enerji Terimi, $E_{\text{üçlü}}^{\psi-\phi}$	55
4.2.3.4 Amino Asit Üçlüsünün Merkezindeki Amino Asitin $C_\alpha - C_\beta$ Kovalent Bağının Üçlüdeki Tüm $C = O$ Kovalent Bağlarına Göre Hizalanmasını λ Dihedral ve $\bar{\theta}$ Bağ Açılarını ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\lambda-\bar{\theta}}$	57
4.2.3.5 Amino Asit Üçlüsünün Merkezindeki Amino Asitin $C = O$ Kovalent Bağının Üçlüdeki Tüm $C_\alpha - C_\beta$ Kovalent Bağlarına Göre Hizalanmasını τ Dihedral ve $\hat{\theta}$ Bağ Açılarını ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\tau-\hat{\theta}}$	58
4.2.3.6 Amino Asit Üçlüsündeki Özdeş Omurga Atomları Arasında Tanımlanan Dört Farklı θ Bağ Açılarını ile Tasvir Edilen Enerji Terimi, $E_{\text{üçlü}}^\theta$	59
4.2.3.7 Amino Asit Üçlüsünün Merkezindeki Amino Asitin Kenar Zincir Atomları Arasında Tanımlanan χ Dihedral Açılarını ile Tasvir Edilen Enerji Terimi, $E_{\text{üçlü}}^\chi$	60
4.2.3.8 Amino Asit Dörtlüleri için Etkileşim Enerji Terimleri, $E_{\text{dörtlü}}^{C_\alpha}$ ve $E_{\text{dörtlü}}^{NCOC_\beta}$	61

4.2.4 Protein Konformasyonu Toplam Enerjisi	63
4.2.5 Monte Carlo Protein Dinamiği Algoritması	65
4.2.6 Monte Carlo Simülasyonundan Elde Edilen Bulgular	67
5. TARTIŞMA ve SONUÇ	71
KAYNAKLAR.....	76
EKLER	87
Moleküler Dinamik Simülasyon Sonuçları	88
Zincir Üzerinde Yerel Olan Amino Asit Üçlüleri için Minimum Enerji Grafikleri ..	93
ÖZGEÇMİŞ	97

KISALTMA ve SİMGELER

NMR	: Nükleer Manyetik Rezonans
PDB	: Protein Data Bankası
LJ	: Lennard Jones
pH	: Hidrojen Kuvveti
$\text{\AA}$	: Ångstrom
F	: Kuvvet
V	: Enerji
T	: Sıcaklık
t	: Zaman
m	: Kütle
v	: Hız
$\dot{v}$	: İvme
γ	: Sürtünme Sabiti
η	: Rastgele Kuvvet
C_α	: α -Karbon atomu
L	: Protein Zincir Uzunluğu
N	: Etkileşme Sayısı
ΔH_{kal}	: Kalorimetrik Entalpi
ΔH_{vH}	: van't Hoff Entalpi
C_P	: Isı Kapasitesi
T_{max}	: Isı Kapasitesinin Maksimum Olduğu Sıcaklık
κ	: Kooperativite Parametresi
k_B	: Boltzmann Sabiti
r	: Mesafe
θ	: Bükülme açısı
ϕ	: Burulma açısı
V_{konf}	: Konformasyonel Enerji
r^0	: Denge Mesafesi
θ^0	: Bükülme Denge Açısı
ϕ^0	: Burulma Denge Açısı
K_r	: Gerilme Terimi Etkileşim Büyüklüğü
K_θ	: Bükülme Terimi Etkileşim Büyüklüğü
K_ϕ	: Burulma Terimi Etkileşim Büyüklüğü
EDP	: Etkileşim Düzen Parametresi
MEDP	: Mutlak Etkileşim Düzen Parametresi
GEDP	: Göreli Düzen Parametresi
UEDP	: Uzun-Erişimli Düzen Parametresi
AEEDP	: Açısal Etkileşimli Etkileşim Düzen Parametresi

OES	: Ortalama Etkileşim Sayısı
MDM	: Moleküler Dinamik Motoru
OIGS	: Ortalama İlk Geçiş Süresi
t_k	: Katlanma Süresi
r_{ese}	: Erim Mesafe Kriteri
Q	: Katlı Yapıya Benzerlik Parametresi

TABLO LİSTESİ

	<u>Sayfa</u>
Tablo 3.1 : Katlı yapılarından bağısız etkileşim ağları elde edilen protein kümesi. ...	22
Tablo 4.1 : Dört donör akseptör çifti setleri için farklı d zincir mesafelerinde gözlemlenen toplam hidrojen bağı sayıları.....	46
Tablo 4.2 : Tüm farklı ailelere mensup proteinlerden, tamamıyla α -yapılı ve tamamıyla β -yapılı proteinlerden oluşturulan üç farklı protein setinde en sık gözlemlenen ilk on amino asit üçlülerinin gösterimi.	49
Tablo A.1 : Basitleştirilmiş zincir modeli ile moleküler dinamik simülasyonu ile incelenen protein içi etkileşim ağlarının simülasyon parametreleri ve çıktıları.	88

ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 2.1 : 2 boyutlu kare latis üzerinde protein katlı yapı temsili.	5
Şekil 2.2 : Protein zinciri üzerinde ardışık olan amino asitler arasında tanımlanan gerilme, bükülme ve burulma terimlerinin gösterimi.	8
Şekil 2.3 : 10-12 Lennard-Jones etkileşme enerjisinin kısa (siyah) ve uzun erişimli (kırmızı) formlarının mesafeye göre değişiminin gösterimi.	12
Şekil 3.1 : GoMDM çalışma şeması; girdileri ve girdilerin GoMDM'nin içindeki işleyişi.	16
Şekil 3.2 : Protein kodu 1ss1 olan protein için farklı etkileşim sona erim mesafe r_{ese} kriterlerine karşılık gelen etkileşim ağlarının gösterimi.	22
Şekil 4.1 : 1ss1 proteini için farklı etkileşim ağlarının gösterimi	24
Şekil 4.2 : Etkileşim sona erim mesafesi ve kooperativite arasındaki ilişki	26
Şekil 4.3 : OES ve kooperativite parametreleri arasındaki ilişki	27
Şekil 4.4 : Deneysel Katlanma oranları ile farklı EDP parametreleri arasındaki ilişki gösterilmesi.	29
Şekil 4.5 : Simülasyon katlanma oranları ile mutlak etkileşim düzen parametresi arasındaki ilişki	30
Şekil 4.6 : Simülasyon katlanma oranları ile Görelî etkileşim düzen parametresi arasındaki ilişki	31
Şekil 4.7 : AEEDP parametresinin simülasyondan elde edilen katlanma oranları ile ilişkisi.	34
Şekil 4.8 : AEEDP parametresi ile deneysel katlanma oranları arasındaki ilişki	34
Şekil 4.9 : Geliştirdiğimiz AEEDP etkileşim düzen parametresi ile termodinamik kooperativite arasındaki ilişkinin gösterimi.	35
Şekil 4.10 : Protein zinciri üzerinde; (a) i 'nci ve j 'inci amino asitler arası öklid mesafesi r ve protein zinciri üzerindeki mesafe tanımının $d = i - j $, ve (b) ardışık amino asit üçlü ve dörtlülerinin notasyonları ve zincir üzerinde gösterimi.....	38
Şekil 4.11 : Alanin amino asitinin β -karbon (${}_AC_\beta$) atomları arasında farklı d zincir mesafelerinde elde edilen etkileşim enerjilerinin öklid mesafesi r_{12} 'ye göre değişimi.....	41
Şekil 4.12 : Alanin amino asitinin omurga azot ${}_AN$ atomu ile omurga oksijen ${}_AO$ atomlarının oluşturdukları (${}_AO_1, {}_AN_2$) ve (${}_AN_1, {}_AO_2$) atom çiftleri için farklı d zincir mesafelerinde elde edilen etkileşim enerjilerinin öklid mesafesi r_{12} 'ye göre değişimleri.....	43
Şekil 4.13 : β -yapılı proteinlerde omurga atomları arasında oluşan asidik $C_\alpha - H \cdots O$ ve $N - H \cdots O$ hidrojen bağları gösterimi.....	44

Şekil 4.14	: Hafif hidrojen atomu tanımlaması içermeyen (a) $N-H \cdots O$ hidrojen bağı oluşum kriterleri ve (b) eğer kriterler yerine getirilmiş ve hidrojen bağı oluşmuşsa ilgili geometrinin kayıt edilmesi ve binlere ayrılabilmesi için tanımlanan dört mesafenin gösterimi.....	44
Şekil 4.15	: Tüm farklı protein katlı yapılarını, sadece tamamıyla α -yapılarını ve sadece tamamıyla β -yapılarını içeren üç farklı protein setinde amino asitlerin gözlemlenme sıklığının gösterimi.....	47
Şekil 4.16	: Tüm farklı protein katlı yapılarını, sadece tamamıyla α -yapılarını ve sadece tamamıyla β -yapılarını içeren üç farklı protein setinde en sık gözlemlenen ilk yüz amino asit üçlüsünde, amino asitlerin gözlemlenme sıklığının gösterimi.....	48
Şekil 4.17	: Protein zinciri üzerinde ardışık olan (a_1, a_2, a_3) amino asit üçlüsündeki, üç amino asit çiftlerinde $C_\alpha - C_\beta$ kovalent bağlarının birbirlerine göre hizalanmalarını tasvir etmek için tanımlanan η_{12} , η_{23} , η_{13} dihedral açıları ve θ_{123} bağ açısı terimlerinin gösterimi.	51
Şekil 4.18	: Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3), \eta_{12}, \eta_{23}, \eta_{13}, \theta_{123})$ enerji fonksiyonunun minimum değere sahip olduğu η_{12} , η_{23} , η_{13} ve θ_{123} parametrelerinin bin kombinasyonlarının gösterimi.	52
Şekil 4.19	: Her (a_1, a_2, a_3) amino asit üçlüsündeki üç amino asit çiftindeki $C=O$ kovalent bağlarının birbirlerine göre hizalanmaları ve geometrik tercihlerini belirlemek için kullanılan ξ_{12} , ξ_{23} , ξ_{13} dihedral açıları ve $\tilde{\theta}_{123}$ eğilme açısı terimlerinin gösterimi.	53
Şekil 4.20	: Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\xi-\tilde{\theta}}((a_1, a_2, a_3), \xi_{12}, \xi_{23}, \xi_{13}, \tilde{\theta}_{123})$ enerji fonksiyonunun minimum değere sahip olduğu ξ_{12} , ξ_{23} , ξ_{13} ve $\tilde{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi	54
Şekil 4.21	: Her (a_1, a_2, a_3) amino asit üçlüsünde tanımlanan ψ_{12} , ϕ_{12} , ψ_{23} ve ϕ_{23} dihedral açılarının grafiksel gösterimi.	55
Şekil 4.22	: Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerji fonksiyonunun minimum değere sahip olduğu ψ_{12} , ϕ_{12} , ψ_{23} ve ϕ_{23} parametrelerinin bin kombinasyonlarının gösterimi.....	56
Şekil 4.23	: (ALA, ALA, ALA), (GLU, ALA, LEU), (GLY, ALA, GLY) ve (THR, ALA, THR) amino asit üçlülerinin ayrı ayrı tüm parametre bin kombinasyonlarının ve karşılık gelen $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerjilerin gösterimi.	57
Şekil 4.24	: Her (a_1, a_2, a_3) amino asit üçlüsünün merkezindeki a_2 amino asitinin $C_\beta^2 - C_\alpha^2$ kovalent bağının üç amino asitteki $C=O$ kovalent bağına göre hizalanmaları ve geometrik tercihlerini belirlemek için kullanılan λ_{21} , λ_{22} , λ_{23} dihedral açıları ve $\bar{\theta}_{123}$ eğilme açısı terimlerinin gösterimi. .	58
Şekil 4.25	: Her (a_1, a_2, a_3) amino asit üçlüsünün merkezindeki a_2 amino asitinin $O=C$ kovalent bağının üç amino asitteki $C_\alpha - C_\beta$ kovalent bağına göre hizalanmaları ve geometrik tercihlerini belirlemek için kullanılan τ_{21} , τ_{22} , τ_{23} dihedral açıları ve $\hat{\theta}_{123}$ bağ açısı terimlerinin gösterimi.	59
Şekil 4.26	: Her (a_1, a_2, a_3) amino asit üçlüsündeki özdeş omurga atomlarının oluşturdukları $\check{\theta}_{123}$, θ_{123} , $\tilde{\theta}_{123}$ ve $\bar{\theta}_{123}$ bağ açısı terimlerinin gösterimi.	60

Şekil 4.27	: (a_1, a_2, a_3) amino asit üçlüsü için pozitif yüklü lizin amino asitinin merkezde olduğu yani $a_2 \equiv \text{LYS}$ seçildiği durumda; χ_2 dihedral açılarının ve bu açılar hangi atom dörtlüleri gözönüne alındığının gösterimi.	61
Şekil 4.28	: Protein zinciri üzerinde ardışık amino asit dörtlülerinin C_α atomlarının geometrik tercihlerini belirlemek için kullanılan $r_{C_\alpha^1 C_\alpha^4}$, mesafesi, $\theta_{123}^{C_\alpha}$ ve $\theta_{234}^{C_\alpha}$ bağ açısı ve $\Omega_{1234}^{C_\alpha}$ dihedral açısı terimlerinin gösterimi.	62
Şekil 4.29	: (a) NMR deneysel yöntemiyle elde edilen 1bdd proteinin orijinal katlı yapısı (macenta renkli) ile simülasyondan elde edilen ve deneyden elde edilen yapıya göre en düşük RMSD değerine sahip protein yapısının (yeşil renkli) hizalanmış hallerinin, (b) gevşetilmiş-katlı yapı (mavi renkli) ve simülasyondan elde edilen katlı yapı (yeşil renkli) hizalanmış hallerinin, (c) (b)'deki yapıların kenar zincir atomları ile birlikte gösterimi.	68
Şekil 4.30	: (a) X-ışını kristalografi deneysel yöntemiyle elde edilen 1enh proteinin orijinal katlı yapısı (macenta renkli yapı) ile simülasyondan elde edilen ve deneyden elde edilen yapıya göre en düşük RMSD değerine sahip protein yapısının (yeşil renkli yapı) hizalanmış halinin, (b) gevşetilmiş-katlı yapı (mavi renkli) ve simülasyondan elde edilen katlı yapısının (yeşil renkli) hizalanmış hallerinin gösterimi.	69
Şekil 4.31	: 1enh proteininin $T_{\text{sim}} = 3.1$ 'de elde edilen E ve RMSD trajektörileri, bu trajektörden elde edilen serbest enerji profili ve ısı kapasitesi değişimi.....	70
Şekil B.1	: Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\lambda-\bar{\theta}}((a_1, a_2, a_3)_i, \lambda_{21}, \lambda_{22}, \lambda_{23}, \bar{\theta}_{123})$ enerji fonksiyonunun minimum değere sahip olduğu $\lambda_{21}, \lambda_{22}, \lambda_{23}$ ve $\bar{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi.....	93
Şekil B.2	: Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\tau-\hat{\theta}}((a_1, a_2, a_3)_i, \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123})$ enerji fonksiyonunun minimum değere sahip olduğu $\tau_{21}, \tau_{22}, \tau_{23}$ ve $\hat{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi	94
Şekil B.3	: Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\theta}((a_1, a_2, a_3)_i, \theta_{123}, \check{\theta}_{123}, \tilde{\theta}_{123}, \bar{\theta}_{123}, \hat{\theta}_{123})$ enerji fonksiyonunun minimum değere sahip olduğu $\theta_{123}, \check{\theta}_{123}, \bar{\theta}_{123}$ ve $\hat{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi	95
Şekil B.4	: Ortanca amino asitin ALA olduğu durumdaki amino asit üçlüleri için $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerji fonksiyonunun minimum değere sahip olduğu $\psi_{12}, \phi_{12}, \psi_{23}$ ve ϕ_{23} parametrelerinin bin kombinasyonlarının gösterimi	96

ÖZET

PROTEİN KATLANMASI DİNAMİĞİ: TRANSFER EDİLEBİLİR ETKİLEŞİM POTANSİYELİ DİZAYNI ve FARKLI KOMPLEKSLİK SEVİYELERİNE SAHİP YENİ MONTE CARLO ve MOLEKÜLER DİNAMİK PROTEİN MODELLERİNİN GELİŞTİRİLMESİ

Gökhan SELAMET

Doktora Tezi, Biyoenformatik ve Bilişimsel Biyoloji Anabilim Dalı

Tez danışmanı: Doç. Dr. Tuğba TAŞKIN TOK

II. Tez danışmanı: Prof. Dr. Hüseyin KAYA

Haziran 2023, 97 sayfa

Protein katlanması, biyolojik bilimlerdeki en zorlu problemlerden bir tanesidir. Sahip oldukları amino asit dizisi, zincir uzunluğu, katlı yapı topolojisi vb. özelliklere bağlı olarak, proteinler katlanma kooperativitesi ve katlanma oranlarında saçılım gibi farklı termodinamik ve kinetik davranışlar sergilerler. Bu davranışların anlaşılması ancak farklı karmaşıklık seviyelerine sahip protein modellerinin ayrıntılı incelenmesi ile gerçekleşebilir. Tezin ilk bölümünde, protein içi etkileşim ağlarının, katlanma termodinamiği ve kinetiği üzerindeki etkileri basitleştirilmiş zincir temsilli içeren moleküler dinamik model yaklaşımı ile 13 farklı protein incelendi. Bariyersiz davranıştan, iki durumlu davranışa termodinamik davranışa geçişin dikkate değer bir şekilde ortalama etkileşim sayısına bağlı olduğu gözlemlendi. Ayrıca, önerilen yeni etkileşim düzen parametresi AEEDP, bağımsız etkileşimlere ilaveten açısız etkileşimleri de içerisine dahil ederek, proteinin hem kinetik hem de termodinamik davranışların açıklanmasında kullanılabildiği görüldü. İkinci bölümde, tüm atomların temsil edildiği Monte Carlo simülasyon yöntemi için transfer edilebilir etkileşim potansiyeli tasarlandı. 25,801 proteinin katlı yapısı üzerinden veri madenciliği kullanarak, (i) atom çiftleri etkileşmesi (ii) $N-H\cdots O$ hidrojen bağları, (iii) 7 farklı ardışık amino asit üçlü terimi, (iv) ardışık aminoasit dörtlüsü için istatistiksel etkileşim potansiyeli geliştirildi. Atom çiftleri etkileşmelerinde, atomların zincir üzerinde ki bir birlerine olan mesafesi incelenerek toplam çift çeşidinin 195,223'e çıkartılması tartışılmıştır. Monte Carlo simülasyon yöntemi ile modellenen iki α -yapılı protein ikisi de başarı ile kararlı yapılarına katlanmış, en düşük 2.29Å RMSD değerine sahip yapı elde edilmiş ve kooperatif katlanma - açılma geçişleri gözlemlenmiştir.

Anahtar Kelimeler: Protein Katlanması, Protein İçi Etkileşim Ağları, Moleküler Dinamik Gō Modeli, Transfer Edilebilir Protein Modelleri, Bütün Atom Temsilli Monte Carlo Protein Modeli

ABSTRACT

PROTEIN FOLDING DYNAMICS: DESIGNING TRANSFERABLE INTERACTION POTENTIALS and DEVELOPING NEW MONTE CARLO and MOLECULAR DYNAMIC PROTEIN MODELS WITH DIFFERENT LEVEL OF COMPLEXITIES

Gökhan SELAMET

Ph.D. Thesis, Bioinformatics and Computational Biology

Advisor: Doç. Dr. Tuğba TAŞKIN TOK

II. Advisor: Prof. Dr. Hüseyin KAYA

June 2023, 97 pages

Protein folding is one of the most challenging problems in biological sciences. Depending on amino acid contents, chain length, native topology, etc., proteins exhibit experimentally different thermodynamic and kinetic behaviors such as the level of cooperativity and diversity in the folding rates. Understanding of these phenomenon requires detailed analysis of theoretical protein models with different level of complexities. In the first part of this thesis, the effects of intra-chain interaction networks on protein folding thermodynamics and kinetics are investigated by coarse-grained molecular dynamic models for 13 different proteins. It was found that the level of thermodynamic cooperativity from downhill to two-state behavior remarkably depends on the average connectivity of the intra-chain interaction network. A newly defined contact order parameter which takes into account angular interaction terms besides the non-bonded interactions (COWAT) clarifies both the kinetic and thermodynamic behaviours. At the second part, new transferable interaction potential is developed for all-atom Monte Carlo protein models. From data mining 25,801 protein folded structures, statistical potential for: (i) two-body interactions, (ii) $N-H\cdots O$ hydrogen bonds, (iii) 7 different local consecutive residue triplets (iv) local consecutive residue quadruples terms are generated. The effect of the chain distance between two atoms on their two-body interaction is further investigated and total unique pair size is increased to 195,223. Two studied α -structured protein is modeled with Monte Carlo simulation are both successfully folded into their native structure, with a lowest RMDS value 2.29Å and cooperative folding-unfolding transition is observed.

Keywords: Protein Folding, Protein Intra-chain Interaction Networks, Molecular Dynamics Gō Model, Transferable protein models, All-Atom Monte-Carlo Protein Models

1. GİRİŞ ve AMAÇ

Biyomoleküllerin dinamiğinin anlaşılması sağlık alanında birçok hastalığın teşhis ve tedavi edilmesi bakımından kritik önem taşımaktadır. Özellikle protein katlanması dinamiği hakkında 90'lı yıllardan bu yana yoğun araştırmalar yapılmasına rağmen; gerek deneysel tekniklerin gerekse bilgisayara dayalı modelleme çalışmalarının yetersizliği nedeniyle hala istenilen seviyede sonuçlar elde edilememiştir [1–3]. Bir proteinin sahip olduğu dizi bilgisi, biyolojik olarak aktif ve kararlı olan katlı yapısı, biyolojik fonksiyonu arasındaki ilişkilerin ortaya çıkarılması; proteinlerin sahip oldukları dizi ve katlı yapı çeşitliliği ve fizyolojik şartlara olan hassaslıklarından dolayı inanılmaz zordur. Proteinlerin gösterdikleri kinetik ve termodinamik davranış çeşitliliği [4, 5], mutasyonlar altında gösterdikleri yapısal ve kararlılık değişimleri [6]; bu biyolojik makro-moleküllerin davranışlarının fizyolojik şartlar altında ne kadar değişkenlik gösterildiğini, diğer bir deyişle hassas ve spesifik olduğunu işaret etmektedir. Bugüne kadar proteinlerin farklı karmaşıklık seviyelerinde temsil edildiği, farklı dinamik algoritmaların kullanıldığı, farklı etkileşim potansiyelleri içeren pek çok bilgisayara dayalı protein modelleri geliştirilmiştir [7–11]. Bu modeller yardımıyla proteinlerin katlanma dinamikleri, dizi-katlı yapı-kararlılık-biyolojik fonksiyonları hakkında çok şey anlaşılabilmiş olmasına rağmen halen ihtiyaç duyulan seviyede değildir. Mesela, dinamik konformasyonel geçişler ihtiva eden ve tüm proteinler için deneylerle birebir uyumlu protein katlı yapıları, katlanma termodinamiği ve kinetiği davranışları elde edilebilen bir protein modeli maalesef hala geliştirilememiştir. Başarılı olarak atfedilen modeller; genelde protein spesifik olan [12, 13] veya kısıtlı sayıda çok küçük proteinler için başarılı sonuçlar verebilen [11, 14] protein modelleridir.

Bu tez çalışmasında; amino asit dizi içeriklerine, zincir uzunluklarına ve içinde bulundukları fizyolojik şartlara göre tamamen farklı kinetik ve termodinamik davranışlar sergileyebilen proteinlerin bu davranışlarının altında yatan temel fiziksel prensiplerin ortaya çıkarılması amaçlanmaktadır. Bu bağlamda; deterministik ve stokastik model yaklaşımlarına dayalı farklı algoritma ve simülasyon yöntemleri kullanılarak ve veri

madenciliğine dayalı transfer edilebilir istatistiksel etkileşim potansiyelleri geliştirilerek; farklı karmaşıklık seviyesine sahip zincir temsilli basitleştirilmiş protein spesifik ve atomistik protein modelleri oluşturulmuştur.

Moleküler dinamik simülasyon yöntemi ve basitleştirilmiş zincir temsili içeren protein spesifik model yaklaşımı kullanarak; protein içi farklı etkileşim ağlarının, protein termodinamiği ve kinetiği üzerinde ki etkisi detaylı bir şekilde incelenmiş ve deneysel olarak kooperatif davranış gösteren proteinlerin bu davranışlarını betimleyen yeni bir etkileşim ağ düzen parametresi geliştirilmiştir. Diğer taraftan, Monte Carlo simülasyon yönteminin kullanıldığı, proteinlerin tüm ağır atomlarının temsili içeren, bu tez kapsamında geliştirilen ve tüm proteinler için kullanılabilecek transfer edilebilir istatistiksel etkileşim potansiyelleri kullanılan, gerçeğe en yakın fakat çok daha yüksek karmaşıklık seviyesine sahip zincir temsilli yeni bir protein modeli geliştirilmiştir. Geliştirilen bu model, yakın gelecekte bu alandaki bir çok problemin anlaşılmasında ve çözümünde büyük katkısı olacağı tahmin edilmektedir.

Özellikle belirtilmesi gerekir ki, bu tez çalışması kapsamında; yeni bilimsel algoritmalar, yeni protein modelleri, yeni transfer edilebilir istatistiksel etkileşim potansiyelleri ve yeni simülasyon yazılımları bir bütünlük içerisinde geliştirilmiştir.

2. GENEL BİLGİLER

2.1 Protein Katlanması

Protein katlanması problemi 60 yılların başlarından miyoglobin proteinin [15, 16] 3-boyutlu kristal yapısının keşfinden bu yana gizemini korumaktadır.

Proteinleri oluşturan atom veya amino asitlerin 3-boyutlu konformasyon uzayında içine girebilecekleri durum sayısı veya serbestlik derecesi olağanüstü yüksek olup, mesela 100 amino asite sahip bir proteinin oluşturduğu konformasyon uzayında içine girilebilir durumların tamamının ziyaret edilebilmesi çok uzun yıllar alacaktır. Fakat aynı büyüklüğe sahip bir proteinin katlanma süreleri farklılık göstermekle birlikte ortalama olarak mili-saniye mertebesinde. Dolayısıyla, bir protein katlanırken konformasyon uzayının bütün durumları ziyaret edilmeksizin, en düşük enerjili katlı yapıya doğru bir seçicilik veya yönlendirme neticesinde çok daha kısa sürede ulaşmakta ve bu Levinthal paradoksu olarak adlandırılmaktadır. [17] Bu bağlamda, farklı amino asit içeriği ve dizilimine sahip proteinlerin deneylerden elde edilen ve en düşük enerjili oldukları kabul edilen 3-boyutlu katlı yapılarına [18]; hangi fiziko-kimyasal süreçler sonucu ulaştığının anlaşılması, protein katlanması çalışmalarının en temel amaçlarından bir tanesidir.

Bir canlıda yüzbinlerce farklı protein, farklı miktarlarda bulunmaktadır. Örnek olarak, insanda 20.687 adet protein kodlayıcı gen bulunur [19], yani insan vücudunda yaklaşık olarak 70 ile 100 bin arası farklı protein bulunduğu düşünülmektedir. Doğada ki milyonlarca farklı canlı türü olduğu düşünüldüğünde toplamda milyarlarca farklı dizilimde protein olduğunu kabaca hesaplayabiliriz. Her proteinin kendine has bir zincir dizilimi ve buna bağlı olarak kristal yapısı bulunmaktadır.

Günümüzde deneysel olarak proteinlerin 3-boyutlu yapıları X-Işını kristalografi [20], NMR [21], elektron mikroskopi [22] vb. yöntemler ile elde edilmektedir. Farklı deneysel yöntemlerle elde edilen protein katlı yapıları; 1971 yılından bu yana herkesin erişimine açık olan protein veri bankasında tutulmaktadır [23].

Bu veri bankası ilk kurulduğunda toplamda sadece 7 adet proteinin kristal yapısı elde edilmişti [23]. Ancak günümüzde bu veri bankası kanalıyla 200,000'e yakın proteinin katlı yapısına ulaşmak mümkün olup, bu sayı her geçen gün hızla artmaya devam etmektedir [24]. Kurulduğu yıldan bu yana, çözümlenen kristal yapı sayısı 28,000 katlık bir artış göstermesine rağmen, doğada bulunan toplam protein çeşitliğinin milyarlar mertebesinde bulunduğunu göz önüne aldığımızda, hala yapısı bilinen protein sayısı oldukça azdır. Bir proteinin katlı yapısının deneysel yöntemler (özellikle, X-ışını kristalografi) kullanılarak elde edilmesi hem maliyetli ve hem de zaman alıcı bir işlemdir. Ayrıca, kristalleştirmekte ki ve sentezlemekte ki zorluklar sebebi ile deneysel olarak her proteinin kristal yapısının elde edilmesi her zaman mümkün olamamaktadır [25].

Katlı veya fonksiyonel yapısı bilinmeyen proteinlerin sayısının fazlalığı ve deneysel zorluklar göz önünde bulundurulduğunda, proteinlerin katlı yapılarının bilişimsel yöntemler kullanılarak ortaya çıkarılması kritik önem taşımaktadır.

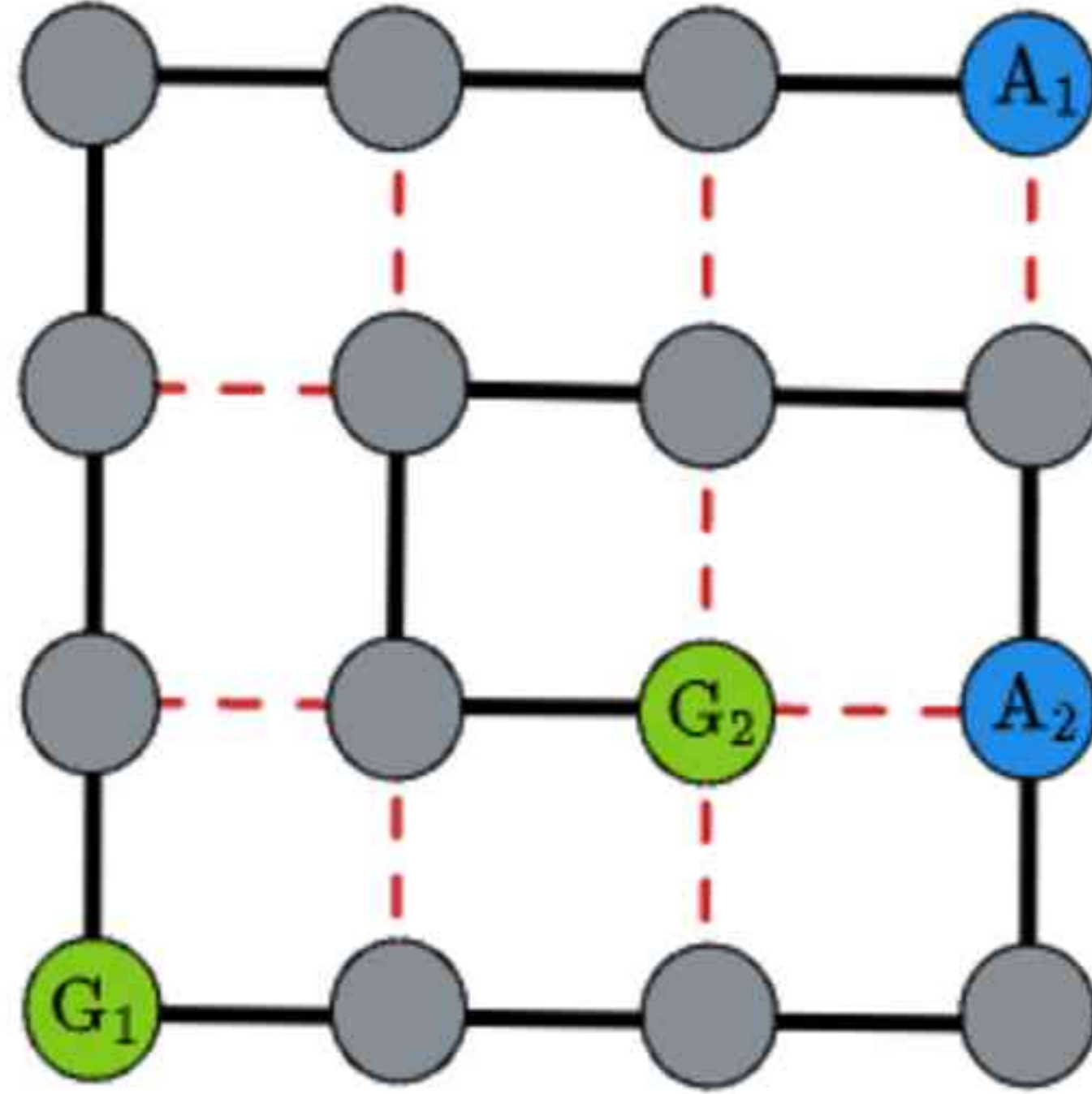
2.2 Protein Katlanma Dinamiğinin Modellenmesi ve Yapı Tahminlerinin Gelişimi

2.3 Protein Spesifik Model Yaklaşımları: Gō Modeli

İlk olarak 1975 yılında Nabuhiro Gō tarafından önerilen bu model yaklaşımı [26–29], protein dinamiği modelleme çalışmalarında özel bir yer tutmaktadır. Gō modeli, ulaşılması hedeflenen protein katlı yapısında bulunan etkileşmelerin, o proteinin dinamiğinin temel belirleyicisi olduğunu varsaymaktadır.

Daha detaylı ifade etmek gerekirse; Gō modeli yaklaşımında sadece ve sadece katlı yapıda birbirleriyle etkileşen atom veya amino asit çiftleri dinamik süresince birbirleriyle etkileşebiliyorlar ve dolayısıyla konformasyon enerjisine katkıda bulunabiliyorlar. Örnek olarak Şekil.2.1'de 2-boyutlu kare latiste katlı yapısı verilen bir protein incelendiğinde; iki tane alanin (A_1 ve A_2) ve iki tane glisin (G_1 ve G_2) amino asitlerinin bulunmakta olduğu ve katlı yapıda sadece (A_2, G_2) amino asit çiftinin etkileştiği görülecektir. Esasen konformasyonel geçişler sırasında bu dört amino asit için altı tane olası amino asit çifti (A_1, A_2), (G_1, G_2), (A_1, G_1), (A_1, G_2), (A_2, G_1), ve (A_2, G_2) etkileşimi mümkün olabilecektir. Örneğimizde olduğu gibi; eğer katlı yapıda sadece (A_2, G_2) amino asit çifti etkileşiyorsa, diğer beş amino asit çifti etkileşim mesafesine gelseler dahi etkileşmedikleri kabul edilmektedir. Diğer bir deyişle, kullanılan enerji fonksiyonu

her proteinin katlı yapısındaki etkileşmelere göre dizayn edilmekte ve sadece o protein için kullanılabilmektedir. Bu nedenle, Gō modeli protein spesifik model olarakta adlandırılmaktadır.



Şekil 2.1 : 2 boyutlu kare latis üzerinde protein katlı yapı temsili. Bu temsili gösterimde gri renkli daireler alanin (ALA, G) ve glisin (GLY, G) olmayan amino asit çeşitlerini, yeşil renkli daireler glisin (G₁ ve G₂) ve mavi renkli daireler alanin (A₁ ve A₂) amino asitlerini, daireler arasındaki düz kalın siyah çizgiler kovalent bağlarını, kesikli kırmızı çizgiler ise katlı yapıdaki etkileşimleri tasvir etmektedir. Katlı yapıda etkileşen tek alanin glisin çifti, (A₂, G₂)’dir.

Ne kadar fiziksel olduğu halen tartışma konusu olan Gō modeli; özellikle deneysel olarak iki-durum termodinamik [30, 31] (keskin sigmoidal veya "ya hep ya hiç" benzeri) ve kinetik davranışı gösteren proteinlerin bu davranışlarının altında yatan mekanizmaların neler olduğunun anlaşılmasında, ve katlanma oranlarındaki çeşitlilik/farklılıkların protein topolojisinin bir temsili olan Etkileşim Düzen Parametresi (EDP) [32] arasındaki ilişkinin gösterilmesindeki başarısından sonra yaygın kullanılmaya başlanmıştır.

Belirtilmesi gerekir ki, bir proteinin dinamik davranışlarını Gō modeli kullanarak inceleyebilmek için o proteinin katlı yapısının bilinmesi gerekmektedir. Atom veya amino asitlerin sadece latis köşelerinde (ayrıklaştırılmış uzayda) olabileceği modellerin dışında [33], sürekli uzayın her noktasında bulunabilecek [34] (yani, çok daha yüksek sayıda olası konformasyon geometrilerinin mümkün olduğu) modellerde atom veya amino asitlerin

birbirlerine göre mesafeleri ve/veya ilgili diğer açı terimlerinin katlı yapıdaki değerleri denge değerleri olarak kullanılmaktadır.

2.3.1 Amino Asitlerin C_α Atomları ile Temsil Edildiği Kabalaştırılmış Protein Moleküler Dinamik Gō Modeli

Zincir temsilli protein modellerinde amino asitler; (i) sadece tek atom (genellikle C_α atomu) ile [34], (ii) omurga atomlarının C_α atomu ve kenar zincir grubunun C_β atomu ile [35], veya (iii) tüm ağır atomları ile temsil edilebilir [36]. Dolayısıyla her protein temsili farklı seviyelerde komplekslik, karmaşıklık ve serbestlik derecesini ihtiva edebilecektir.

Bu tez çalışmasının bir parçası olarak amino asitlerin sadece C_α atomları ile temsil edildiği Gō modeli ele alınacak ve Langevin dinamiği içeren Moleküler Dinamik Simülasyon yöntemi kullanılarak; farklı mesafe kriteri tanımlarına göre proteinlerin katlı yapılarında hiyerarşik bir şekilde oluşturulan etkileşim ağlarının protein termodinamiği ve kinetiğindeki rolü detaylı incelenecektir. Buna ek olarak, α -yapılı, β -yapılı ve $\alpha + \beta$ -yapılı protein ailelerine mensup iki-durum termodinamik ve kinetik davranışı gösteren her bir protein ailesinden 4 tane olmak üzere toplam 12 proteinin bu davranışlarını betimleyebilecek ve ayrıca bu proteinler için oluşturulan farklı etkileşim ağlarından elde edilen katlanma oranlarının hem deneysel gözlemler ile hem de kendi aralarında yüksek korelasyon verecek yeni bir etkileşim düzen parametresi geliştirilecektir.

2.3.2 Model Konformasyonel Enerji Bileşenleri

Amino asitlerin sadece C_α atomları tarafından temsil edildiği en basit zincir temsilli moleküler dinamik Gō protein modelleri, 2000'li yıllarında başlarında kullanılmaya başlanmıştır [37,38]. Bu model yaklaşımında proteinlerin konformasyon enerjisi, V_{konf} ; (i) aralarında kovalent bağ olmayan amino asitler arasında gerçekleşen etkileşimleri tasvir eden ve "*yerel olmayan*" veya "*bağısız*" etkileşim enerjisi olarak adlandırılan $V_{\text{bağısız}}$ şeklinde ifade edilen enerji terimi, ve (ii) protein ikincil yapılarının (sarmallar, β -tabakaları vd.) katlı yapıda gözlemlenen geometrik detaylarda tasvirini sağlamak için protein zinciri üzerinde ardışık olan amino asitler arasında tanımlanan ve "*yerel*" veya "*açısal*" etkileşim enerjisi olarak adlandırılan V_{yerel} veya $V_{\text{açısal}}$ şeklinde gösterilen enerji terimlerinin toplamı (Eşitlik 2.1)

$$V_{\text{konf}} = V_{\text{bağısız}} + V_{\text{açısal}} \quad (2.1)$$

olarak tanımlanmaktadır.

Amino asitler veya atomlar üzerine etkiyen kuvvetler veya kuvvet alanları ise konformasyon enerjisinin negatif gradiyenti olarak hesaplanmaktadır (bkz. Eşitlik 2.2).

$$F = -\nabla V_{\text{konf}} \quad (2.2)$$

Kuvvet alanları, protein katlanması simülasyon çalışmalarında, dinamik süresince birbirleriyle etkileşim içerisinde bulunan parçacıkların nasıl etkileşeceklerinin, fiziksel ve matematiksel olarak tanımlanmasıdır. Bir amino asitin, başka herhangi bir amino asit ile ne koşulda ve ne şiddette etkileşim içerisinde olacağının belirlenmesi kuvvet alanlarının tanımlanması ile gerçekleşir. Bu nedenle, bir modelleme çalışmasında ki en önemli bileşen kuşkusuz tanımlanan kuvvet alanlarıdır. Parçacıkların birbirleri arasındaki etkileşimlerin sonucunda gerçekleşecek konformasyonel geçişleri veya parçacıkların yer değiştirmesini; model için tanımlanan kuvvet alanları belirler. Benzer şekilde, konformasyon uzayının ne derece düzgün ve doğru olarak örneklendirileceği; yine tanımlanan kuvvet alanları tarafından belirlenir. Dolayısıyla, deneysel gözlemlerle uyumlu bir protein katlanması dinamiği ancak kuvvet alanların doğru bir şekilde tasvir edilmesi ile gerçekleşir [39]. Eksik veya yanlış tanımlanan kuvvet alanlarından elde edilecek simülasyon sonuçları deneyler ile örtüşmeyen sonuç üreteceklerdir.

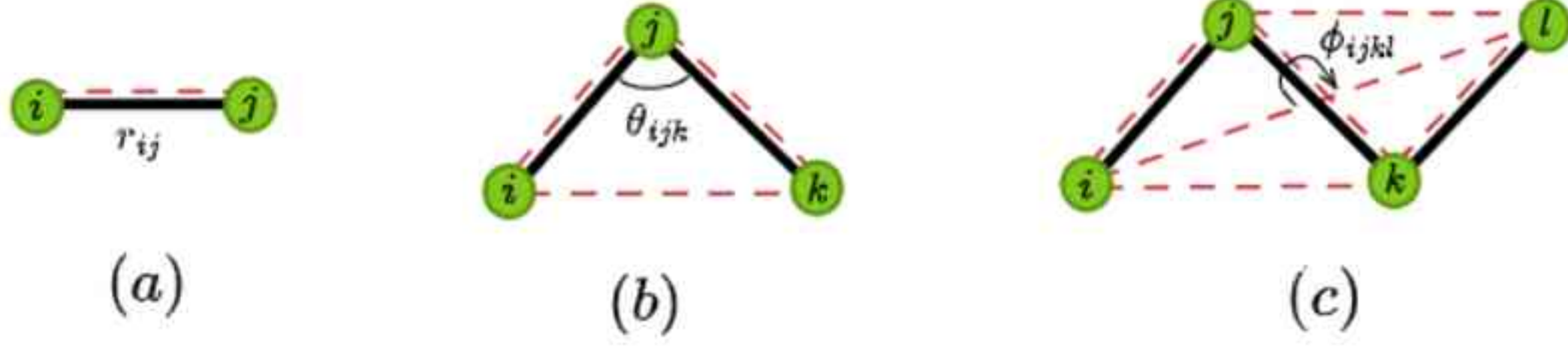
Gö modelinde kullandığımız etkileşim enerji terimlerini yerel ve yerel olmayan etkileşimler olmak üzere iki ayrı gruba ayrılarak aşağıda açıklanacaktır.

2.3.2.1 Yerel - Açısal etkileşimler

Yerel etkileşimler (açısal etkileşimler), zincir üzerinde birbirlerine kovalent bağı ile bağlı olan ardışık atom veya amino asitler arasında tanımlanan etkileşimlerdir. Protein zinciri üzerinde yerel olan bu atom veya amino asit grupları; aralarındaki kovalent bağlarından dolayı daha kısıtlı geometrilere sahip olabilmekte, fakat aynı zamanda protein zinciri üzerinde birbirine yakın (ardışık) oldukları için entropik olarak bu kısıtlı geometrilere daha kısa sürede ulaşabilmektedirler.

Mevcut modelleme çalışmamızda gerilme, bükülme ve burulma olmak üzere üç farklı açısal etkileşim enerji terimi ele alınmıştır [40]. Basitleştirilmiş zincir temsili içeren modelimizde yukarıda ifade edildiği gibi; her bir amino asit sadece bir C_{α} atomu ile temsil

edilmektedir. Bu nedenle, protein katlı yapısından elde edilecek ve simülasyonlarda denge değerleri olarak kullanılacak tüm mesafe ve açı parametrelerinin değerleri, C_α atomları arasında tanımlanmıştır.



Şekil 2.2 : Protein zinciri üzerinde ardışık olan amino asitler arasında tanımlanan (a) gerilme, (b) bükülme ve (c) burulma terimlerinin gösterimi. Yeşil daireler amino asitleri (modelde C_α atomları) ve kalın siyah çizgiler kovalent bağlarını temsil etmektedir. Kırmızı kesik çizgiler ise her açısal terimde birbiri üzerine kuvvet uygulayan amino asit çiftlerini göstermektedir.

Açısal toplam potansiyel enerji; gerilme, burulma ve bükülme enerji terimlerinin toplamına eşittir (Eşitlik 2.3).

$$V_{\text{açısal}} = V_{\text{gerilme}} + V_{\text{bükülme}} + V_{\text{burulma}} \quad (2.3)$$

Her açısal etkileşim enerji teriminin konformasyonel enerjiye katkısı, protein zincir uzunluğuna bağlı olup sabittir. Mesela, zincir uzunluğu (L) olan bir protein için gerilme, bükülme ve burulma enerji terimlerinden kaynaklanan etkileşim sayıları, sırasıyla $L - 1$, $L - 2$ ve $L - 3$ 'tür.

2.3.2.2 Gerilme terimi

Gerilme terimi protein zinciri üzerinde ardışık iki amino asit arasındaki kovalent bağ uzunluğunun, katlı yapıdan elde edilen denge mesafesi değeri etrafında harmonik osilatör gibi salınmasını tanımlamakta ve matematiksel formu Eşitlik 2.4'de verilmektedir.

$$V_{\text{gerilme}} = \sum_{\text{gerilme}}^{N-1} K_r (r_{ij} - r_{ij}^0)^2 \quad (2.4)$$

Eşitlikte 2.4'de (i, j) aminoasit çifti için protein katlı yapısından elde edilen denge mesafesi r_{ij}^0 ile gösterilmiştir. Amino asit çiftinin bulunduğu konformasyondaki mesafesi ise r_{ij} ile gösterilmiştir. Diğer taraftan, $(r_{ij} - r_{ij}^0)^2$ terimindeki parantezin karesi denge mesafesi r_{ij}^0 etrafında harmonik salınımı sağlamak ve r_{ij} 'nin değerini r_{ij}^0 'nin değeri ile aynı veya çok yakın değerlerde kalmasını sağlamaktadır. Yine, Eşitlik 2.4'de K_r terimi

etkileşim büyüklüğünü temsil etmekte ve aksi belirtilmediği sürece, amino asit farklılığı gözetmeksizin değeri 100ϵ olarak sabit tutulmuştur.

2.3.2.3 Bükülme terimi

Bükülme veya eğilme enerji terimi protein zinciri üzerinde ardışık olan üç amino asit arasında tanımlanmaktadır. Daha açık anlatımla, bu ardışık üç amino asiti birbirine bağlayan iki kovalent bağ arasında tanımlanan bükülme açısı θ_{ijk} 'nin (bkz. Şekil 2.2(b)), yine protein katlı yapısından elde edilen θ_{ijk}^0 denge değeri etrafında harmonik osilatör gibi salınımını tasvir etmektedir.

$$V_{\text{bükülme}} = \sum_{\text{bükülme}}^{N-2} K_{\theta} (\theta_{ijk} - \theta_{ijk}^0)^2 \quad (2.5)$$

Bükülme enerji teriminin matematiksel formu Eşitlik 2.5'de verilmiştir. Bu eşitlikte, K_{θ} terimi etkileşim büyüklüğünü temsil etmektedir. Aksi belirtilmedikçe amino asit farklılığı gözetmeksizin etkileşim büyüklüğü $K_{\theta} = 20\epsilon$ olarak alınmış ve sabit tutulmuştur.

2.3.2.4 Burulma Enerji Terimi

Burulma enerji terimi ise protein zinciri üzerinde ardışık olan dört (i, j, k, l) amino asitleri arasında tanımlanmaktadır. Şöyle ki, ilk üç (i, j, k) amino asitlerinin oluşturduğu plaka ile son üç (j, k, l) amino asitlerinin oluşturduğu plaka arasında tanımlanan burulma açısı ϕ_{ijkl} (bkz. Şekil 2.2(c)) ile, yine protein katlı yapısından elde edilen ϕ_{ijkl}^0 denge değeri farkının bir kosinüs fonksiyonu olarak tasvir edildiği ve matematiksel formu Eşitlik 2.6'da verilen etkileşim enerji terimidir.

$$V_{\text{burulma}} = \sum_{\text{burulma}}^{N-3} \{K_{\phi}^{(1)} [1 - \cos(\phi_{ijkl} - \phi_{ijkl}^0)] + K_{\phi}^{(3)} [1 - \cos 3(\phi_{ijkl} - \phi_{ijkl}^0)]\} \quad (2.6)$$

Eşitlik 2.6'da, $K_{\phi}^{(1)}$ ve $K_{\phi}^{(3)}$ terimleri etkileşim büyüklüklerini temsil etmektedir. Aksi belirtilmediği sürece amino asit farklılığı gözetmeksizin etkileşim büyüklüğü sırasıyla $K_{\phi}^{(1)} = 1\epsilon$ ve $K_{\phi}^{(3)} = 0.5\epsilon$ olarak alınmış ve sabit tutulmuştur.

Eşitlik 2.6 ile verilen burulma enerji terimi, periyodik matematiksel fonksiyonlar içermesi dolayısıyla gerilme ve bükülme enerji terimlerinden farklıdır. Burulma enerji terimi periyodu 2π olan $\cos(\phi_{ijkl} - \phi_{ijkl}^0)$ fonksiyonu ve periyodu $\frac{2}{3}\pi$ olan $\cos 3(\phi_{ijkl} - \phi_{ijkl}^0)$ fonksiyonu terimlerini içeren iki kısımdan oluşmaktadır. Bu nedenle, burulma enerji terimindeki serbetlik derecesi, gerilme ve bükülme enerji terimlerine nazaran daha

yüksektir. Belirtilmesi gerekir ki; $x = 0$ olduğunda $\cos x = \cos 3x = 1$ değerlerini almaktadır. Dolayısıyla, $\phi_{ijkl} = \phi_{ijkl}^0$ olduğu zaman Eşitlik 2.6'daki $1 - \cos(\phi_{ijkl} - \phi_{ijkl}^0)$ ve $1 - \cos 3(\phi_{ijkl} - \phi_{ijkl}^0)$ ifadelerinin her ikisi de sıfıra eşit olmaktadır. Diğer bir deyişle, eğer burulma açısı ϕ_{ijkl} 'nin değeri, burulma açısının denge değeri olan ϕ_{ijkl}^0 'ye eşit ise burulma enerji teriminden konformasyon enerjisine herhangi bir katkı gelmemektedir.

Benzer durum, $r_{ij} = r_{ij}^0$ ve $\theta_{ijk} = \theta_{ijk}^0$ olduğu durumlarda sırasıyla gerilme ve bükülme enerji terimleri için de geçerlidir. Buradan anlaşılabileceği üzere; tüm açısal enerji terimlerinin konformasyonel enerjiye katkısı, sadece açısal parametre değerlerinin denge değerlerinden farklı olduğu durumlarda gerçekleşmektedir. Bu katkı, sürekli pozitif değerler içermekte olup, amino asitlere uygulayacakları kuvvetler onların denge değerlerine sahip olmalarını sağlayacak doğrultuda olacaktır. Diğer bir deyişle, kullandığımız Gō modeli için protein katlı yapı konformasyon enerjisine, açısal enerji katkısı sıfırdır ve sadece bağımsız etkileşmeler konformasyonel enerji değerini belirler.

2.3.3 Yerel Olmayan - Bağımsız Etkileşmeler

Bağımsız etkileşimler, protein zincir üzerinde aralarında kovalent bağı bulunmayan amino asit arasında gerçekleşen etkileşimler olarak tanımlanır. Fiziko-kimyasal olarak proteinler Van der Waals etkileşmesi, yük-yük (Coulomb) etkileşmesi, hidrojen bağı, sülfür bağı vd. etkileşmeler protein kararlılığı ve katlanması dinamiğinde önemli rol oynayan bağımsız etkileşimlerdir. Özellikle hidrofobik etkileşimlerin sıcaklığa bağımlılığı, farklı pH'larda histidin amino asitinin farklı karakteristikler göstermesi, farklı denatürant oranlarında etkileşim şiddetlerinin değişimi düşünüldüğünde; proteinlerin fizyolojik şartlardaki değişimlere karşı ne kadar hassas olduğu anlaşılabilecektir.

Fakat protein spesifik Gō modelimizde, amino asit çeşitliliği göz önüne alınmamakta ve olası tüm bağımsız etkileşme şiddetlerinin eşit olduğu kabul edilmektedir. Ayrıca, bağımsız etkileşmelerin protein zinciri üzerinde birbirleri arasındaki mesafe $|i - j| \geq 4$ kriterini karşılayan amino asit çiftleri arasında gerçekleştiği kabul edilmektedir. Çünkü tüm açısal enerji terimleri, zincir üzerindeki mesafeleri $|i - j| < 4$ kriterini sağlayan amino asitler arasında hesaplanmaktadır. Dolayısıyla, $|i - j| < 4$ kriterini sağlayan amino asitler arasında bağımsız etkileşme tanımlayıp konformasyon enerjisine dahil etmek, zincir üzerinde yerel olan $|i - j| < 4$ kriterini sağlayan amino asitler arasındaki etkileşmelerin protein dinamiği içindeki ağırlığını aşırı arttıracak ve zincir üzerinde birbirlerine daha uzak

mesafelerde olan amino asitler arasında gerçekleşen etkileşmelerin dinamiğe katkısını azaltacaktır. Diğer bir deyişle, $|i - j| < 4$ kriterini sağlayan amino asitler arasındaki etkileşmelerden açılal enerji terimleri, $|i - j| \geq 4$ kriterini sağlayan amino asitler arası etkileşmelerden bağısız etkileşmeler sorumlu olacaktadırlar.

Kullandığımız Gō modelinde bağısız etkileşme enerjisi; protein katlı yapısında var olmayan sterik etkileşme enerjisi V_{sterik} ve protein katlı yapısında var olan bağısız etkileşme enerjilerinin $V_{katlı}$ toplamından oluşmaktadır (bkz. Eşitlik 2.7).

$$V_{bağısız} = V_{sterik} + V_{katlı} \quad (2.7)$$

2.3.3.1 Sterik etkileşme enerjisi

Sterik etkileşme enerji terimi, protein katlı yapısında birbirleriyle etkileşmeyen amino asitlerin dinamik konformasyonel geçişler sırasında iç içe geçmelerini engellemek için kullanılmaktadır.

$$V_{sterik} = \sum_{i < j-4}^{sterik} \epsilon^{sterik} \left(\frac{r_{sterik}}{r_{ij}} \right)^{12} \quad (2.8)$$

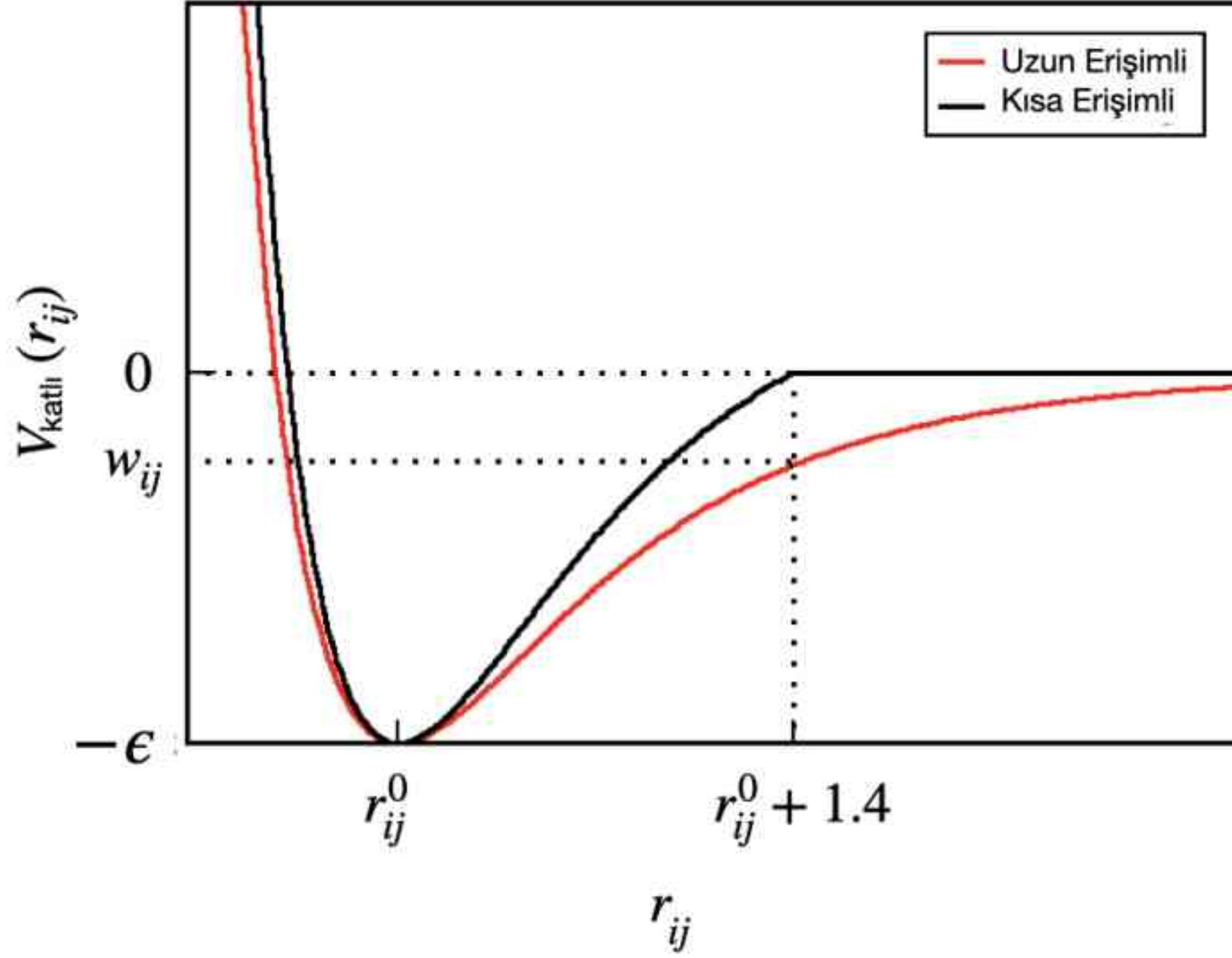
Kullanılan potansiyel enerji formu eşitlik 2.8’de verilmiştir. Eşitlikte kullanılan r_{sterik} bütün amino asit çiftleri için sabit $4\text{\AA}$ alınmıştır. Amino asit çifti (i, j) arasındaki öklidyen uzaysal mesafesi r_{ij} ile gösterilmiştir. Etkileşim şiddeti ϵ^{sterik} ile gösterilmiş, bütün amino asit çiftleri için eşit olarak seçilmiş ve $\epsilon^{sterik} = 1\epsilon$ büyüklüğündedir.

2.3.3.2 Lennard-Jones (LJ) etkileşme enerjisi

Proteinlerin katlı yapısında hangi atom veya amino asitlerin etkileştiği genellikle tanımlanan mesafe kriterlerine göre yapılmaktadır ve literatürde birbirinden oldukça farklı tanımlamalar mevcuttur. Bu mesafe seçiminde hedeflenen şey; iki atomun birbirinden etkilenmeyeceği veya etkisini hissetmeyeceği bir "*etkileşim sona erim mesafesi*" (r_{ese}) tanımlayarak protein katlı yapılarındaki etkileşim ağlarının doğru bir şekilde belirlenmesidir. Basitleştirilmiş model yaklaşımlarında genellikle atomlar arasındaki fiziko-kimyasal farklılıklar göz ardı edilerek tüm farklı atom türleri için aynı r_{ese} mesafe kriteri tanımı yapılmaktadır (mesela, tüm atom çiftleri için $r_{ese} \leq 4\text{\AA}$).

Verilen herhangi bir r_{ese} değeri için katlı yapıda etkileştikleri belirlenen amino asit çiftleri arasındaki bağısız etkileşmeler için literatürde genellikle Eşitlik 2.9 verilen ve mesafeye göre değişimi Şekil 2.3’te kırmızı renkli çizgi ile gösterilen, 10-12 Lennard-Jones (LJ) uzun erişimli etkileşme formu kullanılmıştır.

$$V_{\text{kath}}^{\text{uzun erişimli}}(r_{ij}) = \sum_{i < j-4}^{\text{kath}} \epsilon_{ij}^{\text{kath}} \left[6 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{10} - 5 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} \right] \quad (2.9)$$



Şekil 2.3 : 10-12 Lennard-Jones etkileşme enerjisinin kısa (siyah) ve uzun erişimli (kırmızı) formlarının mesafeye göre değişiminin gösterimi.

Fakat bu tez çalışmasında özellikle matematiksel formu Eşitlik 2.10’da verilen ve mesafeye bağlı değişimi siyah renkli düz çizgi ile Şekil 2.3’te gösterilen kısa erişimli LJ bağımsız etkileşme enerji formu kullanılmıştır. Bunun nedeni ise; 2013 yılında yayınlanmış bir çalışmada [41] deneysel olarak iki-durum termodinamik ve kinetik davranışı gösteren, en küçüğü 41 en büyüğü ise 126 amino asite sahip toplam 52 proteinin kinetiği çalışmasından elde edilen sonuçlardır. Şöyle ki, ilgili proteinlerin deneysel katlanma oranları incelendiğinde; en hızlı ve en yavaş katlanan proteinlerin deneysel katlanma oranlarında $10^{5.96}$ katlık muazzam farklılık olduğu ve benzer $10^{5.81}$ katlık farklılığın sadece kısa erişimli LJ bağımsız etkileşme formunun kullanıldığı model yaklaşımı ile elde edilebilmiş olmasıdır.

$$V_{\text{kath}}^{\text{kısa erişimli}}(r_{ij}) = \sum_{i < j-4}^{\text{kath}} \left(\frac{1}{1 - w_{ij}} \right) \left[\epsilon_{ij}^{\text{kath}} \left(6 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{10} - 5 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} \right) + w_{ij} \right] \quad (2.10)$$

Eşitlik 2.10'daki r_{ij}^0 terimi proteinin katlı yapısından elde edilen (i, j) amino asit çifti için denge mesafesini, r_{ij} terimi konformasyondan konformasyona değeri değişkenlik gösterebilen amino asit çiftleri arasındaki mesafeyi, $\epsilon_{ij}^{\text{kath}}$ terimi ise katlı yapıda etkileşen amino asitlerin bağımsız etkileşim şiddetini ifade etmektedir. $\epsilon_{ij}^{\text{kath}}$ 'nın değeri, amino asit çiftleri arasındaki bütün bağımsız etkileşimler için aynı $\epsilon_{ij}^{\text{kath}} = 1\epsilon$ olarak belirlenmiştir. Ayrıca w_{ij} terimi ise r_{ij}^0 denge değerinden 1.4 Å uzaklıkta kısa erişimli LJ enerji fonksiyonunu sıfıra kaydırmak için tanımlanan ve değeri uzun erişimli LJ enerji formundan elde edilen ve fiziksel bir anlamı olmayan bir kaydırma parametresidir.

2.4 Proteinlerde termodinamik kooperativite

İki durum davranışı gösteren küçük proteinlerin, kalorimetri deneylerinde kalorimetrik entalpi ile van't Hoff entalpi arasındaki oranı κ 'nın 1 civarı değerler aldığı deneysel olarak gözlemlenmiştir [42].

$$\kappa = \frac{\Delta H_{\text{kal}}}{\Delta H_{\text{vH}}} \approx 1 \quad (2.11)$$

Kalorimetrik entalpi ΔH_{kal} , sabit basınç altında farklı sıcaklıklarda ölçülen ısı kapasitesi değerlerinin oluşturduğu, ısı kapasitesi eğrisinin altında kalan alan olarak tanımlanmaktadır. Eşitlik 2.12'de kalorimetrik entalpi için matematiksel formülasyonu gösterilmiştir.

$$\Delta H_{\text{kal}} = \int_{T_0}^{T_1} dT C_P(T) \quad (2.12)$$

Eşitlik 2.12'deki T_0 ve T_1 sıcaklık terimleri, katlanma ve açılma geçişlerinin gerçekleştiği sırasıyla en düşük ve en yüksek sıcaklık değerlerine karşılık gelmektedir.

Geçiş sıcaklığında tanımlanan van't Hoff entalpi ΔH_{vH} ise, ısı kapasitesinin maksimumunun bir fonksiyonu olarak tanımlanmış ve matematiksel formu Eşitlik 2.13'de verilmiştir.

$$\Delta H_{\text{vH}} = 2T_{\text{max}} \sqrt{k_B C_P T_{\text{max}}} \quad (2.13)$$

2.5 Protein katlanma kinetiği

Proteinlerin zamana bağlı davranışları arasında; özellikle katlanma oranlarındaki değişkenlikler (pico saniye mertebesinde [43] - mili saniye mertebesine [44]) , farklı denaturant miktarlarında gözlemlenen lineer olan ve lineerlikten sapan katlanma ve

açılma oranı değişimleri (Chevron grafikleri) hem deneysel [6, 31] hem de teorik [5, 34] çalışmalarda üzerinde en yoğun araştırma yapılan protein kinetik analizleridir. Proteinlerin açık yapıdan katlı yapıya ilk ulaştığı/geçtiği süre veya katlanma süresi (t_k) olarak tanımlanmaktadır. Diğer taraftan, farklı başlangıç koşulları ile elde edilen katlanma sürelerinin ortalaması, "*ortalama ilk geçiş süresi*" (*OIGS*) olarak tanımlanmaktadır (bkz. Eşitlik 2.14).

$$OIGS = \frac{1}{N} \sum_{i=1}^N t_k^i \quad (2.14)$$

Ortalama ilk geçiş süresinin tersi olarak katlanma oranları (k_f) tanımlanmıştır.

$$k_f = \frac{1}{OIGS} \quad (2.15)$$

3. GEREÇ ve YÖNTEM

3.1 Moleküler Dinamik Simülasyon Programı

Proteinlerin modellenmesi çalışmalarında sık kullanılan yöntemlerden birisi moleküler dinamik simülasyon yöntemidir. Özetle bu yöntem sistemdeki parçacıklara etki eden kuvvetlerin hesaplanması ve bu kuvvetlerin neticesinde birim zaman aralığında parçacıkların konumlarının hesaplanarak hareket ettirilmesine dayanmaktadır.

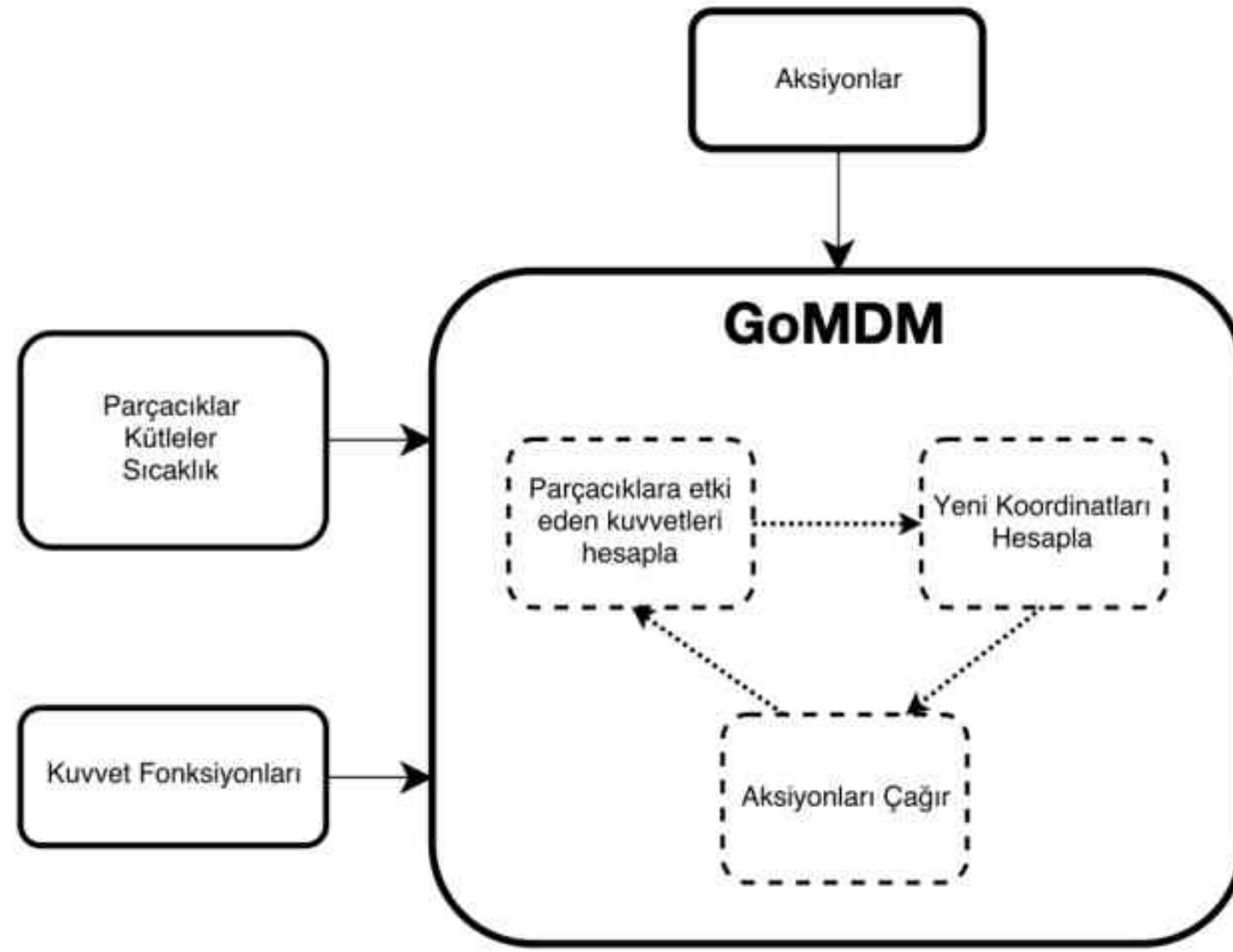
Bu tez çalışması kapsamında geliştirilecek yeni protein modellerinin çalışılması için bir simülasyon programına ihtiyaç vardır. Literatürde protein katlanması için kullanılan bir çok simülasyon programı mevcuttur. Bunlara örnek olarak; GROMACS [45], NAMD [46], LAMPPS [47] verilebilir. Ayrıca protein katlanması içinde kullanılabilecek, çok amaçlı geliştirilmiş moleküler dinamik kütüphaneleri de mevcuttur [48].

Tez kapsamında mevcut olan programların kullanılması değerlendirilmiştir. Ancak iki temel sebepten bu programların kullanılmasından vazgeçilmiştir. Birinci sebep, mevcut programlarda ki kuvvet ve enerji fonksiyonlarının matematiksel formlarının değiştirilmesinin mümkün olmaması yada kişiselleştirmenin kısıtlı olmasıdır. Ayrıca bir çok program mimarisi gereği, kişiselleştirildiği durumda hesaplama performansı düşük olmaktadır. İkinci sebep ise; benzer şekilde simülasyon dinamiği süresince yapılması planlanan analizlerin sisteme dahil edilmesinin çok kısıtlayıcı olmasından kaynaklıdır. Geçmiş çalışmalardan elde edilen tecrübeler, özellikle etkileşim potansiyeli geliştirilmesine yönelik bilimsel çalışmalarda modüler ve esnek simülasyon motorlarının önemiyetinin anlaşılmasını sağlamıştır. Bu noktada, tez çalışması sırasında geliştirilecek moleküler dinamik modeller için yeni bir simülasyon programı tasarlanmıştır.

3.1.1 Moleküler Dinamik Motoru: GoMDM

GoMDM, bu tez çalışmasının bir parçası olarak basitleştirilmiş zincir temsili protein katlanması modelleri için özel olarak tasarlanmış bir moleküler dinamik simülasyon motorudur. Her ne kadar programlama mimarisi, bütün ağır atomların temsil edildiği

bir model için uyumlu olarak tasarlansada, bu tez kapsamında sadece moleküler dinamik simülasyon yöntemi ile sadece basitleştirilmiş zincir temsilli, protein spesifik modeller kullanıldığı için GoMDM'in odak noktası bu yönde daraltılmıştır. GoMDM'nin içerisinde parçacık hareketleri Langevin dinamiği [49] ve Verlet hız algoritması [50] kullanılarak yapılmaktadır. Simülasyon dinamikleri 2003 yılında Kaya ve Chan [40] tarafından yayınlanan makale ile uyumlu olarak, verlet hız algoritması ise Veitshans ve arkadaşlarının [51] makalesini baz alarak yapılmıştır.



Şekil 3.1 : GoMDM çalışma şeması; girdileri ve girdilerin GoMDM'nin içindeki işleyişi.

3.1.2 GoMDM'nin Kullanılması

GoMDM C programlama dilinde yazılmıştır ve kullanılması oldukça kolaydır. Kullanıcı dört adımda simülasyon başlatabilir. Bunlar sırasıyla aşağıdaki listedeki gibidir.

1. Parçacıkları tanımlama
2. Kuvvetleri tanımlama
3. Aksiyonları tanımlama
4. Simülasyonu başlatma

İlk aşamada kullanıcı GoMDM'e parçacıkların başlangıç koordinatlarını, parçacıkların kütlelerini, toplam parçacık sayısını ve simülasyon sıcaklığını vermelidir. İkinci aşamada, kullanıcı hangi kuvvetlerin hangi parçacıklara uygulanacağını belirten fonksiyonları vermelidir. Aynı anda birden fazla kuvvet fonksiyonu GoMDM'de tanımlanabilir.

Tanımlanan her kuvvet fonksiyonunun çıktısı, o kuvvet fonksiyonu tarafından parçacıklara etki eden toplam kuvveti saklayan bir dizide toplanır. Üçüncü aşamada kullanıcı simülasyon süresince yapılacak aksiyonları tanımlar. Tanımlanacak aksiyonlar her bir simülasyon adımından sonra GoMDM tarafından çağrılır. Simülasyon adımının dosyaya kaydedilmesi, kontakların hesaplanması, trajektörü dosyasının kaydedilmesi kullanılabilecek aksiyonlara örnektir. Son adım olarak simülasyonun çalıştırılması kalmıştır. Kullanıcı simülasyon zaman adımı cinsinden simülasyon süresi vererek simülasyonu başlatabilir.

Yukarıdaki adımları içeren örnek kod parçası aşağıda verilmiştir.

```
void main ()
{
    vec3 *particles = ...
    float *mass = ...
    int length = ...

    smtr_init(particles , mass , length , 0.5);

    scg_add_streching_force();
    scg_add_bending_force(20.0);
    scg_add_torsion_force(1.0 , 0.5);
    scg_add_nonbond_force();

    smtr_subscribe_event(print_event , NULL);

    smtr_run_loop(100000);
}
```

3.1.3 GoMDM'in İşlevi

Esasen GoMDM proteinlerin basitleştirilmiş zincir temsili modelinde moleküler dinamik simülasyon çalıştırmak için yazılmıştır. Ancak bölüm 3.1.2'de anlatıldığı ve şekil 3.1'de görülebileceği gibi parçacıklar, kuvvetler ve aksiyonlar kullanıcı tarafından GoMDM'e dışarıdan verilmektedir. Bu durumda kullanıcılar için farklı karmaşıklık seviyelerinde protein yada diğer biyolojik moleküllerin modellenmesinin önü açılmıştır.

$$m\dot{v}(t) = F_{\text{konf}}(t) - m\gamma v(t) + \eta(t) \quad (3.1)$$

GoMDM'nin işlevi parçacıkların yeni koordinatlarını hesaplamak ve verilen aksiyonları çağırmasıdır. GoMDM Langevin dinamiğini kullanır ve hareket denklemi Eşitlik 3.1'de

verilmiştir. Eşitlik 3.1'deki $m, v, \dot{v}, F_{konf}, \gamma, \eta$ sembolleri sırasıyla, kütle, hız, ivme, konformasyon kuvveti, sürtünme sabiti(viscosity) ve rastgele kuvveti ifade etmektedir.

Rastgele kuvvet (η) Eşitlik 3.2 kullanılarak hesaplanır ve Eşitlik 3.3'de verilen otokorelasyon fonksiyonu sağlar.

$$\eta_i = \left(\frac{2m\gamma k_B T}{\delta t} \right)^{\frac{1}{2}} \xi_i \quad (3.2)$$

Eşitlikteki k_B terimi Boltzmann sabitini, T terimi sıcaklığı ve ξ_i terimi ise ortalaması 0 standart sapması 1 olacak şekilde Gaussian dağılımına uyan rastgele sayı üreticisinden alınan sayıdır.

$$\langle \eta(t)\eta(t') \rangle = 2m\gamma k_B T \delta(t - t') \quad (3.3)$$

GoMDM simülasyonun ilk adımında parçacıklara rastgele bir başlangıç hızı tanımlar. Bu hız Eşitlik 3.4'de verilen formüle göre tanımlanır.

$$v_i = \left(\frac{k_B T}{m} \right)^{\frac{1}{2}} \xi_i \quad (3.4)$$

3.1.4 Parçacıklar arası mesafe hesabı

Bir moleküler dinamik simülasyonunda en zaman alıcı hesaplardan bir tanesi parçacıklar arası mesafenin hesaplanmasıdır. Simülasyonlarda kullanılan potansiyel enerji ve kuvvet alanları parçacıkların birbirine olan mesafesine dayalıdır. Farklı kuvvetlerin tekrar tekrar simülasyon içerisinde aynı mesafeleri hesaplamasını önlemek için GoMDM mesafe hesabını kullanıcıya bırakmamıştır. Her simülasyon adımında parçacıklar arasında mesafeleri bir kere ve hızlı şekilde hesaplanır ve kullanıcının erişimine sunar. Böylelikle dinamikte ki kuvvet sayısı ne kadar fazla olursa olsun, hesaplanacak mesafe sayısı sabit tutulmuş olacaktır. Böylelikle GoMDM ile geliştirilecek farklı protein modellerinde, kullanıcı parçacıklar arası konumların hesaplanma optimizasyonu ile ilgilenmeyerek sadece geliştireceği etkileşim potansiyeline odaklanması sağlanmıştır.

Parçacık sayısının N olduğu bir sistemde, her bir simülasyon adımında hesaplanması gereken mesafe sayısı $\frac{1}{2} \times N \times (N - 1)$ kadardır. Yani parçacık sayısının artması hesaplanacak mesafe sayısını üstel olarak arttırmaktadır. Litaratürde bu sorunu çözmeye dayalı, geliştirilmiş farklı algoritmalar mevcuttur [50, 52]. GoMDM'de ise verlet liste algoritması [50] kullanılmaktadır.

Verlet liste algoritması moleküler dinamikte ki parçacıkların hareketinin birim zaman içerisinde kısıtlı yer değiştirme yapabileceğinden yararlanarak kurgulanmıştır. Kısaca açıklamak gerekirse, simülasyon koşullarına göre her parçacığın bir simülasyon adımında alabileceği maksimum mesafe belirlenebilir. Mesela birbirlerine 100 Å uzaklıkta olan iki parçacık bir sonraki simülasyon adımında birbirlerine 10 Å kadarlık yakına o simülasyon koşulunda gelemeye bilirler. Eğer bu iki parçacık 10 Å'den daha yakın mesafelerden itibaren etkileşiyorsa, bu zaman adımında aralarındaki mesafeye hesaplamaya biliriz. Çünkü bir önceki hesapladığımız mesafeye göre biliyoruz ki aralarında ki yeni mesafe etkileşme mesafesinden çok daha fazla olacaktır. Bu durumlar özel tanımlanmış modeller için uygulanabilir. Örneğin bu tez çalışmasında kullandığımız kısa erişimli LJ etkileşim formu için kullanılmaktadır.

Verlet liste algoritması her adımda mesafesi hesaplanacak parçacıkların bir listesi çıkartılır. Bu liste belirlenen bir mesafe kriterinden birbirlerine daha yakın olan parçacıkların listesidir. Yine belirlenen simülasyon adımında sadece listedeki parçacık çiftleri arasındaki mesafeler hesaplanır ve güncelenir. Listedeki parçacıkları hangilerinin olduğunu her simülasyon adımında güncellenmez, daha geniş periyotlarda güncellenir. Böylelikle her adımda $\frac{1}{2} \times N \times (N - 1)$ kadar mesafe hesaplamak yerine sadece verlet listesinin uzunluğu kadar mesafe hesaplar. Sadece listenin güncellenmesi sırasında bütün parçacıklar arası mesafeyi hesaplayarak simülasyonun daha performanslı çalışması sağlanır.

3.1.5 pdb Okuyucu: cpdb

GoMDM sadece moleküler dinamik simülasyon motorudur. Parçacıklar Şekil 3.1'de görülebildiği gibi dışarıdan kullanıcı tarafından verilmektedir. Bu bağlamda parçacıkların başlangıç kordinatlarını ve türleri GoMDM'ye hazırlamak için pdb formatı okuyucu (pdb parser) kütüphanesi yazılmış ve cpdb adı verilmiştir. Pdb dosya formatı proteinlerin deneysel olarak elde edilen kristal yapı bilgisini içeren bir dosya formatıdır [53]. Protein içinde ki her bir aminoasitin ve atomun çeşidi, zincir bilgisi, koordinat vb. bilgileri içeren metin tipinde bir formattır. Moleküler dinamik moturu GoMDM C dili ile yazıldığı için aynı çatı altında simülasyonların çalıştırılması amaçlandığından, cpdb de C dili ile yazılmıştır.

cpdb'nin data yapısı biopython kütüphanesinden esinlenerek yapılmıştır [54]. Data yapısı *pdb* → *zincir* → *aminoasit* → *atom* şeklindedir. Tez kapsamında ki çalışmalarda model bilgisi kullanılmadığı için data yapısına model bilgisi eklenmemiştir.

cpdb'nin mimarisi alternatif atomları da okuyacak şekilde tasarlanmıştır. Ancak mevcut çalışmalarımızda alternatif atom koordinatları kullanılmadığı için cpdb alternatif atomları elemine edecek fonksiyonlar eklenmiştir.

3.1.6 cpdb Kullanımı

cpdb kütüphanesinin kullanımı oldukça basittir. Öncelikle dataların kayıtlı tutulacağı bir pdb tipinde yapı oluşturmamız gerekir. Sonra sadece okuyacağımız pdb dosyasının adresini vererek kristal yapı dosyasında ki bütün bilgileri okuyabiliriz. Son adım olarak işlemiz bittiğinde oluşturduğumuz yapıyı hafızadan sileceğiz. Bu paragrafta anlatılan adımları içeren kod bloğu aşağıda örnek olarak verilmiştir.

```
void main() {  
    pdb *P;  
    P = initPDB();  
    parsePDB (".//1cqu.pdb", P, "");  
    printPDB (P);  
    freePDB (P);  
}
```

Bu yapıda görebildiği gibi cpdb, aynı kod diliminde birden fazla pdb dosyası okumak için uygun yapıda tasarlanmıştır. cpdb'nin asıl amacı data yapılarının GoMDM ile uyumlu olarak tasarlanmasıdır. Böylelikle GoMDM ile entegre bir şekilde çalışabilmektedir.

3.1.7 Kodların erişimi

Bu tez çalışmasında geliştirilen GoMDM ve cpdb kütüphaneleri açık kaynak kodlu olarak Gaziantep Üniversitesi internet sunucularında erişime açılmıştır. Aşağıda ki bağlantıyı kullanarak kodlara ulaşabilirsiniz.

<http://gul3.bim.gantep.edu.tr/~selamet/kod>

3.2 Protein İçi Etkileşim Ağları

Protein katlanması modelleme çalışmalarında, zincir içi etkileşim ağının protein termodinamiği ve kinetiği üzerinde etkilerinin anlaşılması, çözülmesi gereken sorulardan

bir tanesidir. Farklı etkileşim ağlarının, protein davranışı üzerine etkisinin anlaşılması ve bu ağların proteinin termodinamik kooperativitesine ve katlanma kinetiğine etkisinin incelenmesi, aynı protein modelinde etkileşim ağının rolünün anlaşılması; doğru protein modellerinin geliştirilmesine yardımcı olacaktır.

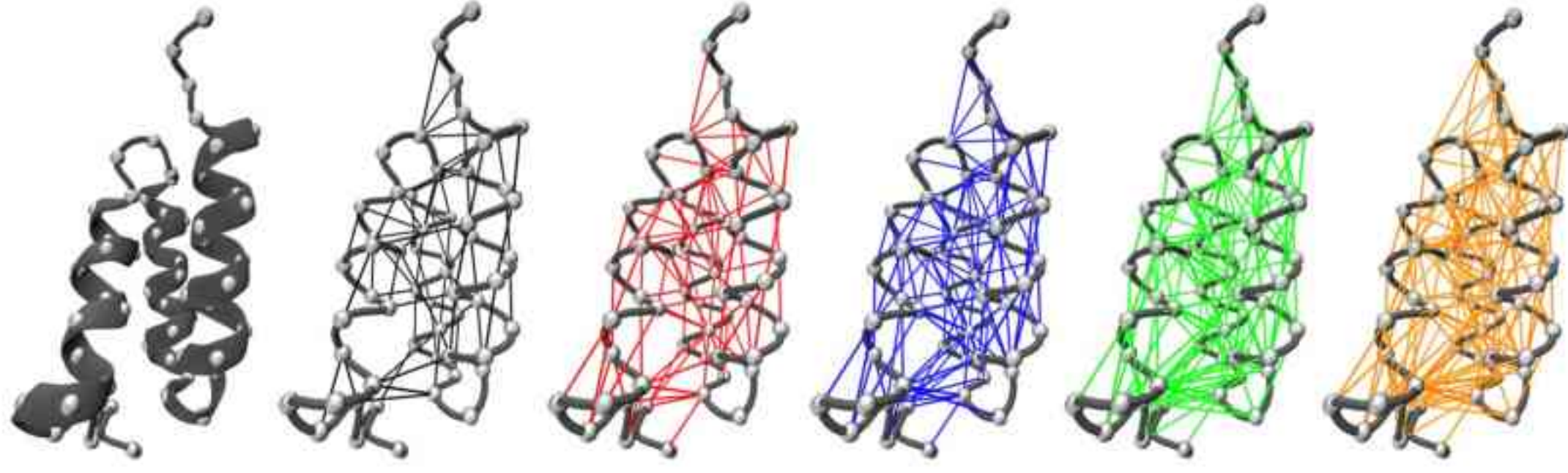
Transfer edilebilir etkileşim potansiyelleri içeren protein modellerinin aksine, Gō modeli gibi protein spesifik modeller, protein içi etkileşim ağlarının incelenmesine olanak sağlamaktadır. Gō modelinde katlı yapıda bulunan bağırsız etkileşmeler ile bulunmayan bağırsız etkileşmeler birbirinden ayrılması aynı protein topolojisinde ve aynı protein modelinde etkileşim ağlarının direk etkisinin çalışmasını kolaylaştırmaktadır.

3.2.1 Protein içi etkileşim ağlarının oluşturulması

Katlı yapıda bulunan etkileşmeler, etkileşim sona erim mesafe kriteri r_{ese} 'ye bağlı olarak belirlenmektedir. Bunun için protein zinciri üzerindeki mesafeleri $i < j - 4$ kriterini sağlayan amino asit çiftlerinin içerdikleri tüm atomları arasındaki mesafelerin r_{ese} 'den küçük olup olmadığı kontrol ediliyor ve eğer iki amino asitin sadece birer atomu arasındaki mesafe r_{ese} mesafesinden daha yakın olsa dahi, o iki amino asitin etkileştiği kabul edilmektedir. Bu şekilde elde edilen etkileşim ağlarındaki denge mesafeleri oldukça yüksek değişkenlik göstermektedir. Çünkü, triptofan, fenilalanin, tirozin gibi büyük amino asitlerin kenar zincirlerinin son atomları etkileşme kriterini sağlarken, bu etkileşen amino asitlerin omurga atomları (dolayısıyla, C_α atomları) birbirlerinden çok uzakta olabilmektedirler. Genel olarak, protein etkileşim ağındaki etkileşen amino asitler arasındaki denge mesafeleri $4 \leq r_{ij}^0 \leq 6 \text{ \AA}$ aralığında gözlemlenmekte, fakat farklı proteinlerin katlı yapılarında $r_{ij}^0 > 12 \text{ \AA}$ olduğu denge mesafeleri sıklıkla olmasa da gözlemlenmektedir.

Bu şekilde, farklı r_{ese} değerleri için aynı protein katlı yapısından farklı bağırsız etkileşim ağları elde etmek mümkündür. Esasen bu şekilde elde edilen etkileşim ağları birbirlerinden tamamen bağımsız olmayıp, r_{ese} mesafe kriteri değeri arttırıldığında bir önceki etkileşim ağına etkileşim ilavesi yapılmış olunmaktadır. Dolayısıyla hiyerarşik şekilde oluşturulmuş etkileşim ağları söz konusudur.

Örnek olarak pdb kodu 1ss1 olan "Immunoglobulin G binding protein A" proteini [55] için oluşturulmuş beş farklı etkileşim ağı Şekil 3.2'de verilmiştir. Etkileşim sona erim mesafe



Şekil 3.2 : Protein kodu 1ss1 olan protein için farklı etkileşim sona erim mesafe r_{ese} kriterlerine karşılık gelen etkileşim ağlarının gösterimi. Sırasıyla siyah, kırmızı, mavi, yeşil ve turuncu renkleri; $r_{ese} = 3.5\text{\AA}$, $4.5\text{\AA}$, $5.0\text{\AA}$, $5.75\text{\AA}$, ve $6.25\text{\AA}$ değerlerinde elde edilen etkileşim ağlarını göstermektedir.

Tablo 3.1 : Katlı yapılarından bağımsız etkileşim ağları elde edilen protein kümesi.

pdb kodu	Protein ailesi	Protein ismi
1bb1	α -yapılı	E3 Binding Domain [56]
1c9o	β -yapılı	Bacillus Caldolyticus Bc-Csp [57]
1cqu	$\alpha + \beta$ -yapılı	50s Ribosomal Protein L9 [58]
1fex	α -yapılı	Trf2-Interacting Telomeric Rap1 Protein [59]
1idy	α -yapılı	Mouse C-Myb Dna-Binding Domain Repeat 3 [60]
1jo8	β -yapılı	Actin Binding Protein [61]
1o6x	$\alpha + \beta$ -yapılı	Procarboxypeptidase A2 [62]
1rlq	β -yapılı	SH3 domain [63]
1shf	β -yapılı	SH3 domain in human fly [64]
1ss1	α -yapılı	Immunoglobulin G binding protein A [65]
1w4j	α -yapılı	Pyruvate Dehydrogenase E2 [66]
2ci2	$\alpha + \beta$ -yapılı	Chymotrypsin Inhibitor 2 [67]
2ptl	$\alpha + \beta$ -yapılı	Protein L [68]

kriteri $3.4 \leq r_{ese} \leq 6.25 \text{\AA}$ aralığında sistematik olarak artırılarak, aynı protein katlı yapısı için bağımsız etkileşimlerin sistematik olarak nasıl arttığı Şekil 3.2’de görülebilmektedir.

3.2.2 Moleküler dinamik simülasyonlarında kullanılan protein kümesi

Protein içi etkileşim ağlarının incelenmesinde, 12 farklı protein seçilmiştir. Bu protein kümesi, dört adeti α -yapılı, dört adeti β -yapılı, ve dört adeti $\alpha + \beta$ yapılı olmak üzere üç farklı protein ailesinden kooperatif iki durum davranışı gösteren proteinlerden oluşmaktadır. Ayrıca bir adet bariyersiz (downhill) davranışı gösteren, α -yapılı protein incelenmiştir. Kullanılan protein kümesinde ki, proteinlerin, pdb kodları, yapı bilgileri detaylarını Tablo 3.1’de detaylı olarak verilmiştir.

4. BULGULAR

4.1 Protein İçi Etkileşim Ağlarının Protein Katlanmasına Etkileri

Bu kısımda, aynı protein yapılarından elde edilmiş farklı etkileşim ağlarının katlanma termodinamiğine ve kinetiğine etkisi incelenmiştir. Farklı etkileşim ağları, Bölüm 3.2.2’de verilen protein kümesi için Bölüm 3.2.1’de detaylı olarak anlatılan sistematik ile oluşturulmuştur. Her bir protein için etkileşim sona erim mesafesi $r_{ese} = 3.25 \text{ Å}$ ’den başlayarak, sistematik olarak 6.4 Å ’e kadar artırılmıştır. Böylelikle bir protein için ortalama 13 tane farklı etkileşim ağı olmak üzere toplamda 212 farklı etkileşim ağı için ısı kapasitesi, serbest enerji profili analizleri yapılmış ve katlanma oranları bulunmuştur. Tez çalışmasında ele alınan proteinler için çalışılan etkileşim sona erim mesafeleri ve bunlara karşılık gelen etkileşim ağları detaylı olarak ekler (A) kısmında ki tabloda verilmiştir.

Çalışılan proteinlerin farklı etkileşim ağları için öncelikle, geçiş sıcaklıkları elde edilmiştir. Geçiş sıcaklığında elde edilen simülasyonlardan serbest enerji profilleri hesaplanmış ve ısı kapasiteleri çizdirilmiş ve κ kooperativite değerleri hesaplanmıştır. Ayrıca proteinlerin katlanma kinetikleri çalışmış, her bir etkileşim ağı için katlanma oranı (k_f) katlanma/açılma geçiş sıcaklığında hesaplanmıştır.

4.1.1 Protein İçi Etkileşim Ağlarının Protein Katlanma Termodinamiğine Etkisi

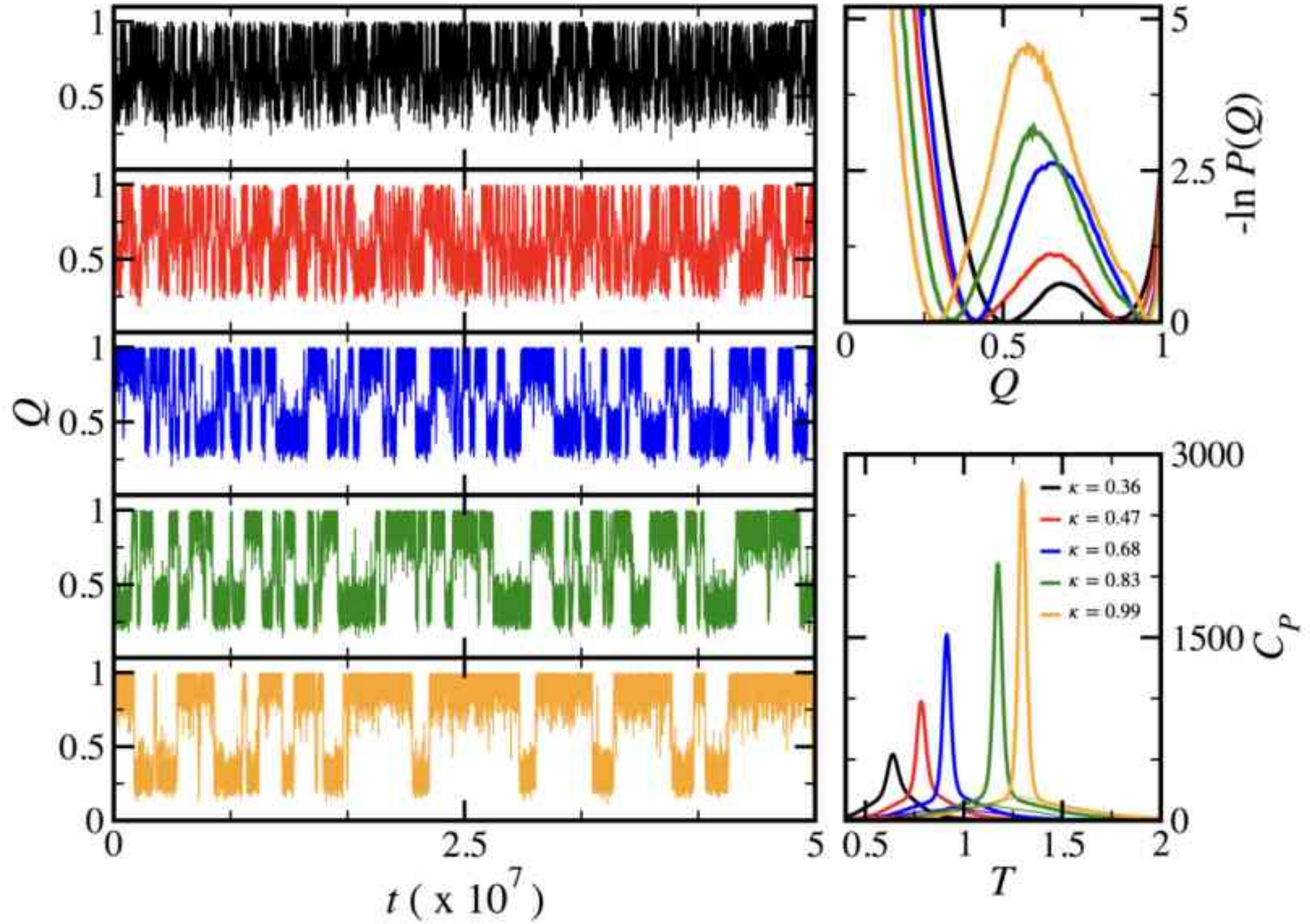
Bu kısımda, örnek olarak seçilen tamamıyla α yapılı Issl proteininden elde edilen farklı etkileşim ağları için geçiş sıcaklıklarında elde edilen trajektori, serbest enerji profili ve ısı kapasitesi grafikleri Şekil 4.1’de verilmiştir. Şekilde ki siyah, kırmızı, mavi, yeşil ve turuncu renkli çizgiler sırasıyla etkileşim sona erim mesafesinin $r_{ese} = 3.5 \text{ Å}$, 4.5 Å , 5.0 Å , 5.75 Å , ve 6.25 Å , değerlerine karşılık gelen protein içi etkileşim ağları için elde edilmiş simülasyon sonuçlarını göstermektedir. Bu protein etkileşim ağları temsili olarak aynı renklendirme ile Şekil 3.2’de de kullanılmıştır.

Şekil 4.1’de kullanılan Q terimi; bir konformasyonda bulunan katlı yapı etkileşme sayısı $N_{bağsız}^{konf}$ ’nin, katlı yapıdaki toplam etkileşme sayısı $N_{bağsız}^{katlı}$ ’ye oranı olarak tanımlanmakta

(bkz. Eşitlik 4.1) ve 0 ile 1 arasında değerler almaktadır. $Q = 1$ değeri protein katlı yapısındaki tüm etkileşmelerin gerçekleştiğini, dolayısıyla katlı yapıyı temsil etmektedir.

$$Q = \frac{N_{\text{bağısız}}^{\text{konf}}}{N_{\text{bağısız}}^{\text{katlı}}} \quad (4.1)$$

Diğer taraftan, $-\ln P(Q)$ serbest enerji profilini ve $P(Q)$ ise simülasyon boyunca gözlemlenen Q değerine sahip konformasyonların normalize edilmiş popülasyonunu ifade etmektedir.



Şekil 4.1 : Protein kodu 1ss1 olan protein için farklı etkileşim sona erim mesafe kriterilerine karşılık gelen trajektorilerin (sol), serbest enerji profillerinin (sağ-üst) ve ısı kapasitelerinin (sağ-alt) gösterimi. Sırasıyla siyah, kırmızı, mavi, yeşil ve turuncu renkleri; etkileşim sona erim mesafesinin $r_{\text{ese}} = 3.5\text{\AA}$, $4.5\text{\AA}$, $5.0\text{\AA}$, $5.75\text{\AA}$, ve $6.25\text{\AA}$ değerlerinden elde edilen etkileşim ağlarına ait simülasyon sonuçlarını göstermektedir.

Trajektori grafiğine bakıldığında (bkz. Şekil 4.1) kolaylıkla gözlenebileceği üzere, etkileşim sona erim mesafesi r_{ese} artırıldıkça katlanma ve açılma süreleri artmaktadır. Diğer bir tabir ile katlanma ve açılma zorlaşmaktadır. Buna bağlı olarak, serbest enerji profilleri incelendiğinde; etkileşim sona erim mesafesinin değerinin artırılmasıyla, geçiş durumu popülasyonunun azaldığı gözlemlenmektedir. Böylelikle açık yapıdan katlı yapıya giderken (veya tersi de doğru) aşılması gereken serbest enerji bariyer yüksekliği

artmakta ve katlı yapı topluluğu ile açık yapı topluluğunun birbirinden daha fazla ayrıştığı sonucu ortaya çıkmaktadır.

Şekil 4.1’de ısı kapasitesinin sıcaklığa bağlı değişimi incelendiğinde, etkileşim sona erim mesafesi r_{ese} arttıkça, geçiş sıcaklıklarının da artmakta olduğu gözlemlenmektedir. Ayrıca geçiş sıcaklığı civarında ısı kapasitesinin aldığı maksimum değerin arttığı ve ısı kapasitesindeki standard sapmanın azaldığı görülmektedir. Şekil 4.1’de beş farklı etkileşim sona erim mesafesi r_{ese} değeri için κ değerleri ayrıca verilmiştir. r_{ese} ’nin değeri 3.25Å’den 6.25Å’a doğru artarken, κ değerlerinin 0.36’dan 0.99’a doğru arttığı gözlemlenmektedir. Başka bir deyişle, simülasyon sonuçlarına göre, 1ss1 proteini $r_{ese} = 3.5\text{Å}$, 4.5Å , 5.0Å , 5.75Å değerlerinde elde edilen etkileşim ağırları için termodinamik olarak kooperatif olmayan bir katlanma/açılma davranışı sergilerken, $r_{ese} = 6.25\text{Å}$ değerinde elde edilen etkileşim ağı için deneysel sonuçlarla uyumlu termodinamik iki durum kooperatif katlanma/açılma davranışı elde edilmektedir.

Daha öncede belirttiğimiz gibi, incelediğimiz 1ss1 proteinin deneysel olarak iki-durum davranışı gösterdiği ve bu davranışı gösteren proteinlerin kalorimetri deneylerinde ölçülen κ değerinin 1 civarında olduğu bilinmektedir. Aynı protein topolojisi, aynı protein zincir temsili ve aynı etkileşim kuvvet alanları kullanılarak yapılan modelleme çalışmalarında dahi; tek başına protein içi etkileşim ağırlarının ne derece düzgün tanımlanıp tanımlanmadığı veya deneysel sonuçlarla uyumlu davranışlar gözlemlmek için etkileşim ağırlarının ne kadar hassas seçilmesi gerektiği elde ettiğimiz sonuçlar ışığında çok daha net ortaya çıkmıştır.

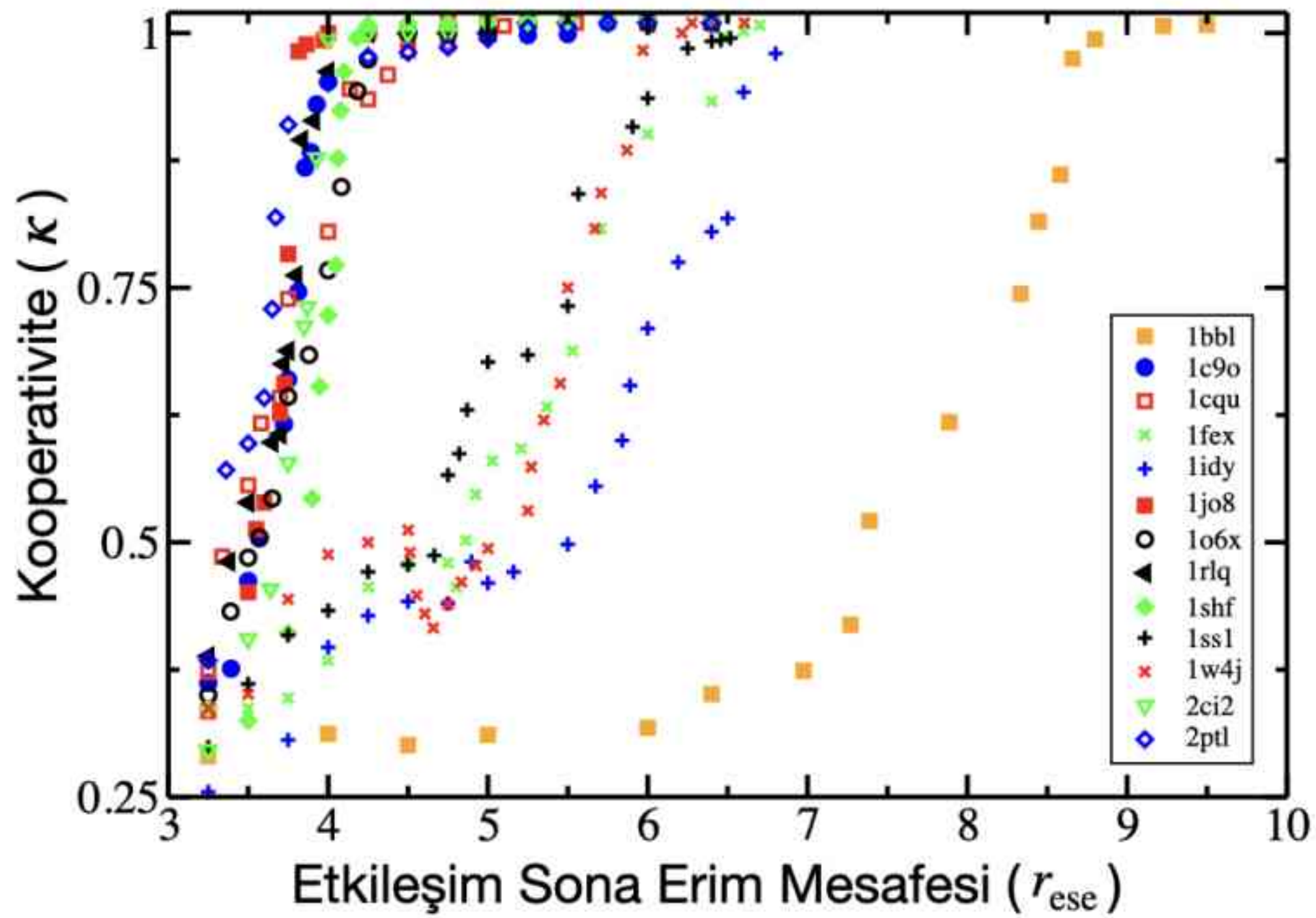
4.1.1.1 Etkileşim Sona Erim Mesafe Kriteri ve Termodinamik Kooperativite İlişkisi

Bir önceki bölümde, örnek olarak ele aldığımız 1ss1 pdb kodlu protein için beş farklı etkileşim sona erim mesafe r_{ese} değerleri için elde edilen etkileşim ağırlarının gösterdiği termodinamik davranışları detaylı olarak incelemiştik. Bu kısımda ise protein kümemizde bulunan 12 proteinin her biri için 15 farklı etkileşim sona erim mesafe r_{ese} değeri kullanılarak elde edilen etkileşim ağırlarının; termodinamik kooperativite davranışına etkilerini topluca inceleyeceğiz. Aynı analiz, deneysel olarak bariyersiz davranış gösteren ve pdb kodu 1bbl olan protein için 30 farklı etkileşim ağı oluşturularak gerçekleştirilmiştir.

Şekil 4.2’de protein kümemizdeki tüm proteinler için çalışılan etkileşim ağırlarının elde edildiği r_{ese} değerlerine karşılık gelen ve bu etkileşim ağırları için hesaplanan termodinamik

kooperativite parametresi κ 'nın deęerleri verilmiřtir. G r ld ę   zere, etkileřim sona erim mesafesi r_{ese} yeterince arttıęında, b t n proteinler deneylerle uyumlu termodinamik kooperativite davranıřı ($\kappa \approx 1$) sergilemeye bařlamaktadır.

Yine řekil 4.2 dikkatli incelendięinde; farklı ailelere mensup proteinlerin farklı r_{ese} deęer aralıklarında k melendikleri g r lecektir. Tamamıyla β ve $\alpha + \beta$ yapılı proteinlerden oluřan ilk grup $r_{ese} \approx 4 \text{  }$ civarında kooperatif katlanma davranıřı g stermeye bařlarken, tamamıyla α yapılı proteinlerden oluřan ikinci grup, $r_{ese} \approx 6 \text{  }$ civarında aynı davranıřı g stermeye bařlamaktadır. Dięer taraftan, deneysel olarak bariyersiz davranıř g steren 1bbl proteini ancak $r_{ese} \approx 8.7 \text{  }$ civarında kooperatif davranıř g sterebilmektedir ki; r_{ese} 'nin bu deęeri fiziko-kimyasal olarak kabul edilebilir sınırların  ok  zerindedir. Esasen bu sonu  modelimizin 1bbl proteini i in deneysel sonu larla  rt řen sonu  verdięini ayrıca iřaret etmektedir.



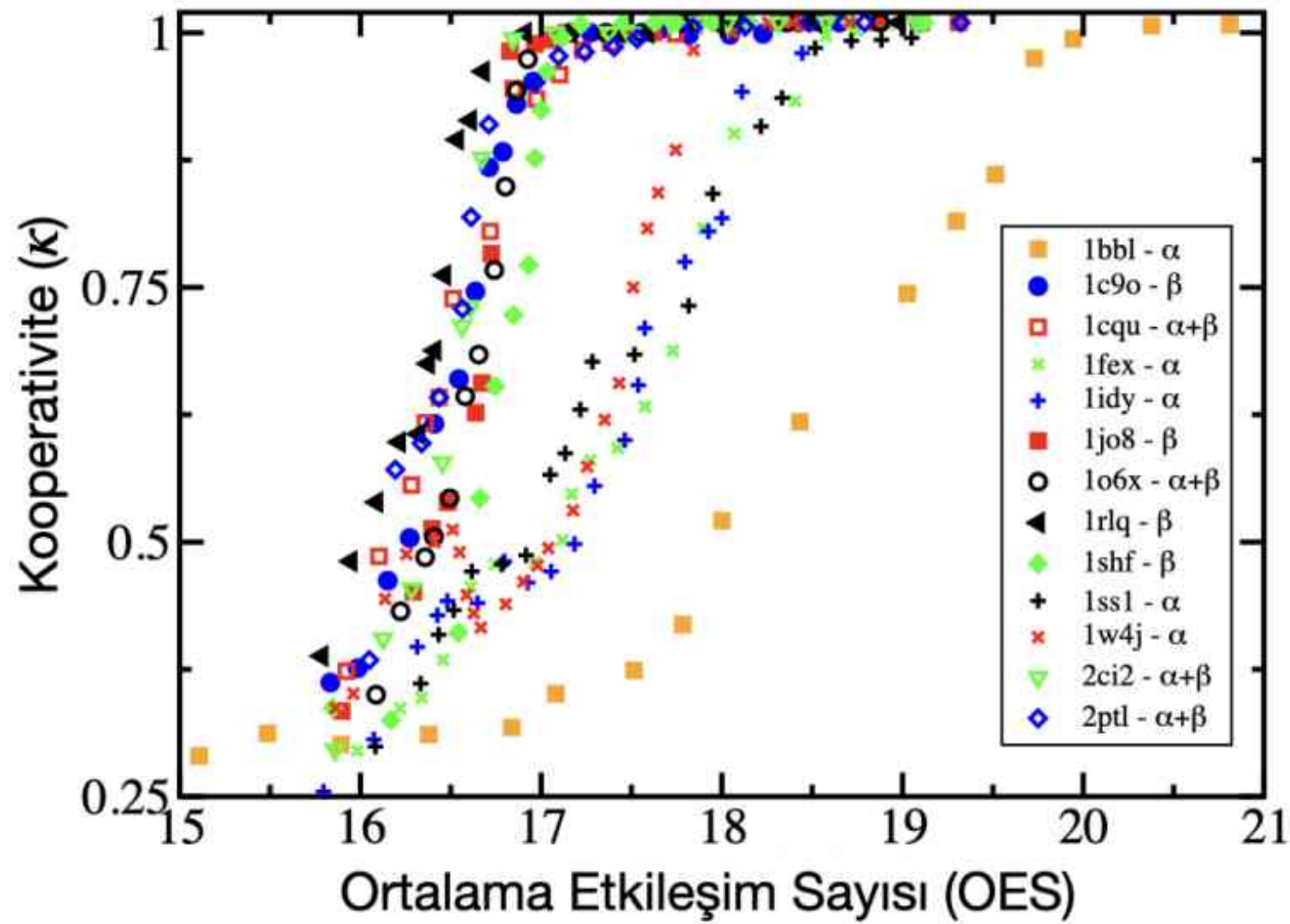
řekil 4.2 : Etkileřim sona erim mesafesi r_{ese} ile r_{ese} 'nin verilen deęerleri kullanılarak oluřturulan etkileřim aęları i in elde edilen kooperativite deęerleri arasındaki iliřkinin g sterimi.

Bu sonu lar ıřıęında, farklı ailelere mensup proteinlerin aynı r_{ese} deęerlerinde elde edilen etkileřim aęları kıyaslandıęında; tamamıyla α yapılı proteinlerin  ok daha d ř k etkileřim sayılarına sahip oldukları bulunmuřtur. r_{ese} deęeri yeterince artınca, etkileřim sayısındaki bu farklılık beraberinde ortadan kalkmaktadır.

4.1.1.2 Ortalama etkileşim sayısı ve termodinamik kooperativite ilişkisi

Etkileşim sona erim mesafesi r_{ese} 'nin değeri artırıldığında, etkileşim ağına yeni bağımsız etkileşimler ilave edilmektedir. Yani etkileşim ağları birbirinden tamamen bağımsız değildir. Aksine, aralarındaki farklılık sadece eklenen yeni etkileşimlerdir. Protein zincir uzunluğu (L) sabit olduğu için etkileşim ağındaki etkileşim veya kontak sayısı arttıkça, amino asit başına düşen ortalama etkileşim sayısı (OES) artmaktadır.

$$OES = \frac{N_{\text{kontak toplam}}^{\text{kontak}}}{L} = \frac{N_{\text{kontak bağımsız}}^{\text{kontak}} + N_{\text{kontak açılal}}^{\text{kontak}}}{L} \quad (4.2)$$



Şekil 4.3 : 12 farklı proteini için amino asit başına düşen ortalama etkileşim sayısı (OES) ile termodinamik kooperativite arasındaki ilişkinin gösterimi.

Eşitlik 4.2'deki $N_{\text{kontak bağımsız}}^{\text{kontak}}$ terimi r_{ese} 'nin verilen değeri ile protein katlı yapısından elde edilen etkileşim ağındaki toplam bağımsız etkileşme sayısını ve $N_{\text{kontak açılal}}^{\text{kontak}}$ terimi ise gerilme, bükülme ve burulma terimlerinden gelen (ve her proteinin tüm etkileşim ağlarında sabit olan) toplam etkileşim sayısını temsil etmektedir. Bükülme ve burulma enerji terimleri çoklu parçacık etkileşimleri olduğu ve her parçacık tüm parçacıklar üzerine kuvvet uyguladığı için bu terimlerden gelecek etkileşim sayısı katkısı; içerdikleri parçacık çifti sayısı kadar olacaktır. Daha açık ifadeyle, her bir bükülme teriminden 3 ve burulma teriminden 6

tane etkileşen parçacık çifti katkısı gelecektir. Bu terimlerdeki etkileşen parçacık çiftileri Şekil 2.2’de kırmızı kesikli çizgiler ile gösterilmiştir. Bir proteinin tüm etkileşim ağlarında $N_{\text{aşısal}}^{\text{kontak}} = 10L - 25$ ’tir ve bu değer sabittir. Dolayısıyla, Eşitlik 4.2’de farklı etkileşim ağları için değişken olan tek parametre $N_{\text{bağısız}}^{\text{kontak}}$ ’dır.

OES ile kooperativite parametreleri arasındaki ilişki incelendiğinde (bkz. Şekil 4.3); etkileşim sona erim mesafeleri için gözlemlenen kümelenmelere benzer kümelenmeler (bkz. Şekil 4.2) göze çarpmaktadır. Protein kümeleri arasındaki ayrışıklık, etkileşme sona erim mesafelerinden elde edilen sonuçlara nazaran bir miktar azalmış olduğu görülmektedir. Her etkileşim ağı için hesaplanan amino asit başına düşen ortalama etkileşim sayısı OES, her bir amino asitin diğer amino asitlerle ne derece etkileştiğinin veya bağlantısı olduğunun ortalama ölçüsü olup, özellikle küçük r_{ese} değerlerinde amino asitlerin etkileşim sayılarında yüksek standart sapmalara sebebiyet verebilmektedir. Diğer bir deyişle, kimi amino asitler hiç bağısız etkileşme yapmazken kimileri çok yüksek sayıda (özellikle protein katlı yapısı merkezinde bulunan amino asitler) bağısız etkileşme yapabilmektedir. Standart sapma r_{ese} değeri arttıkça azalmaktadır.

4.1.2 Etkileşim ağlarının protein katlanma kinetiğine etkisi

Bu bölümde, farklı protein içi etkileşim ağlarının, proteinlerin katlanma kinetiğine olan etkisi incelenmiştir. Katlanma kinetiği simülasyonları her bir protein için elde edilen ısı kapasitesinin maksimum olduğu sıcaklıkta (T_{max}) veya diğer adıyla katlanma/açılma geçiş sıcaklıklarında çalıştırılmıştır.

Bölüm 4.1.1’de 1ss1 proteini için etkileşim sona erim mesafesi arttıkça, katlanma ve açılma geçişlerinin yavaşladığından bahsetmiştik. Bu gözlem genel olarak incelediğimiz bütün proteinler için geçerlidir. Her protein için etkileşim sona erim mesafesi r_{ese} ’nin değeri 3.25Å’den 6.4Å’e arttırıldığı durumda, aynı proteinden elde edilen en yavaş ve en hızlı katlanma oranları arasında 10 tabanlı logaritma ölçeğine göre 1.25 ila 4.17 arasında değişen farklılık gözlemlenmektedir. Katlanma oranlarında en yüksek farklılık gösteren 2ptl proteini için elde edilen 4.17 değerinin anlamı; $r_{\text{ese}} = 3.25\text{Å}$ ve 6.4Å için oluşturulan iki etkileşim ağından elde edilen en yavaş ve en hızlı ortalama katlanma süreleri arasında $10^{4.17} \approx 14,800$ kat farklılık olduğudur.

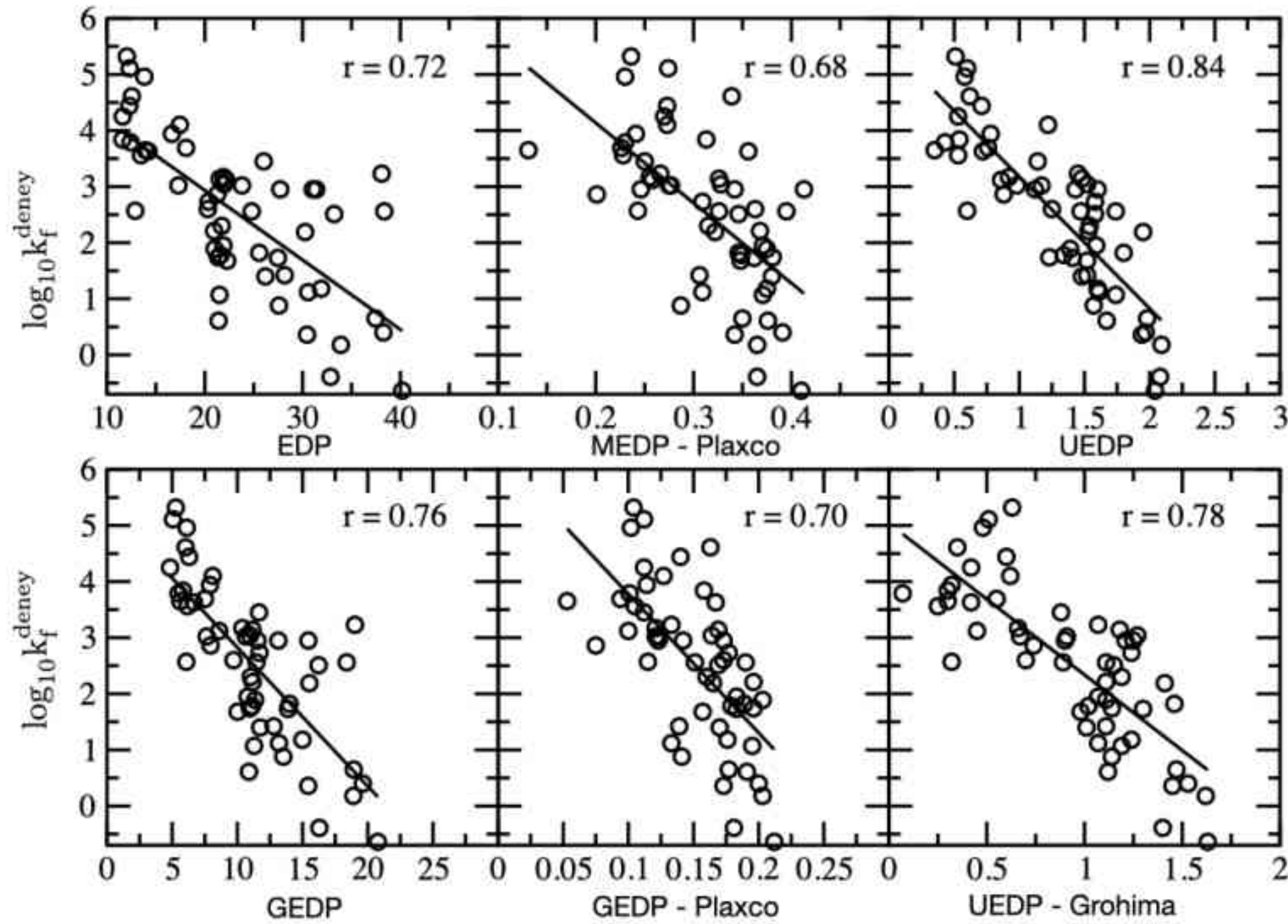
Aynı protein için tanımlanan etkileşim ağlarının, katlanma oranlarında bu kadar yüksek farklılıklara sebebiyet vermesi; protein içi etkileşim ağlarının protein kinetiğinin en temel

belirleyicisi olduğunu ifade etmektedir. İncelediğimiz bütün proteinlerin, tüm etkileşim ağları için simülasyonlardan elde edilen katlanma oranları Ekler A kısmında tablolar halinde detaylı olarak verilmiştir.

4.1.2.1 Literatürdeki etkileşim düzen parametreleri ve katlanma oranları ilişkisi

Etkileşim düzen parametresi (EDP) ilk olarak 1998 yılında Plaxco ve ark. tarafından ortaya atılmıştır [32]. İlgili çalışmada etkileşim sona erim mesafesi $r_{ese} = 6.0 \text{ Å}$ alınarak proteinlerin katlı yapısında etkileştikleri gözlemlenen amino asit çiftlerinin, protein zinciri üzerindeki mesafelerinin toplamının, toplam etkileşme sayısı ile normalize edilmesine dayanan ve protein topolojisini dolaylı bir şekilde betimleyen bir etkileşim düzen parametresi önerilmiştir.

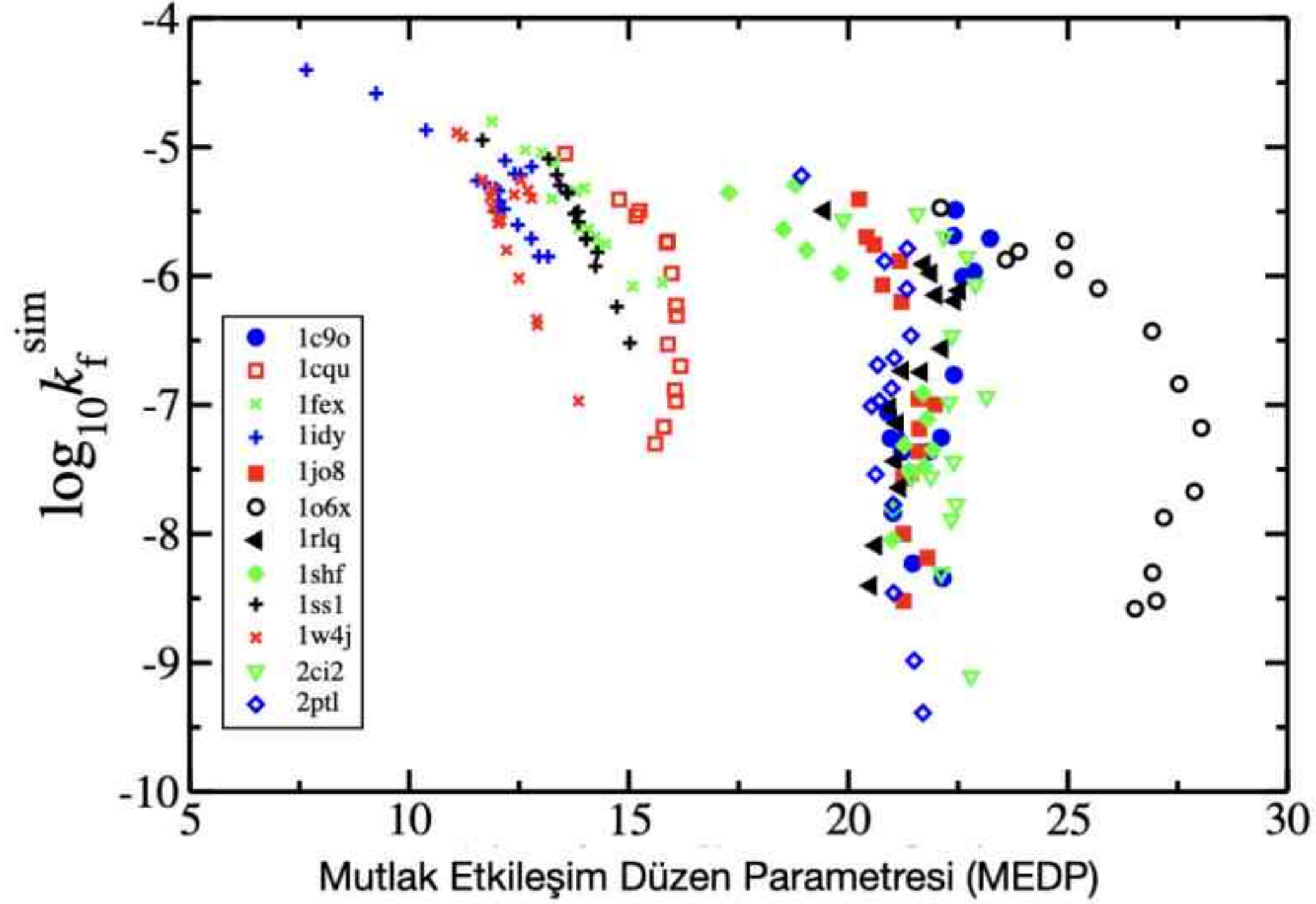
$$MEDP = \frac{1}{N_{\text{bağısız}}^{\text{katlı}}} \sum \Delta Z_{ij} \quad (4.3)$$



Şekil 4.4 : Deneysel katlanma oranları ile farklı EDP parametrelerini arasındaki ilişkilerin gösterimi. Her bir EDP parametresi tanımı için elde edilen Pearson korelasyon katsayısı (r) şeklin içinde verilmiştir.

Eşitlik 4.3'te matematiksel formu verilen etkileşim düzen parametresi, "*mutlak etkileşim düzen parametresi*" (MEDP) olarak adlandırılmaktadır. Eşitlikteki ΔZ_{ij} terimi katlı yapıda bağısız etkileşme yapan amino asitlerin zincir üzerinde birbirlerine

olan mesafelerini (yani, $\Delta Z_{ij} = |i - j|$) ve $N_{\text{bağırsız}}^{\text{katlı}}$ terimi katlı yapıdaki toplam bağırsız etkileşim sayısını temsil etmektedir. Eşitlikteki toplam sadece katlı yapıda bağırsız etkileşme yapan amino asit çiftleri üzerinden yapılmaktadır.



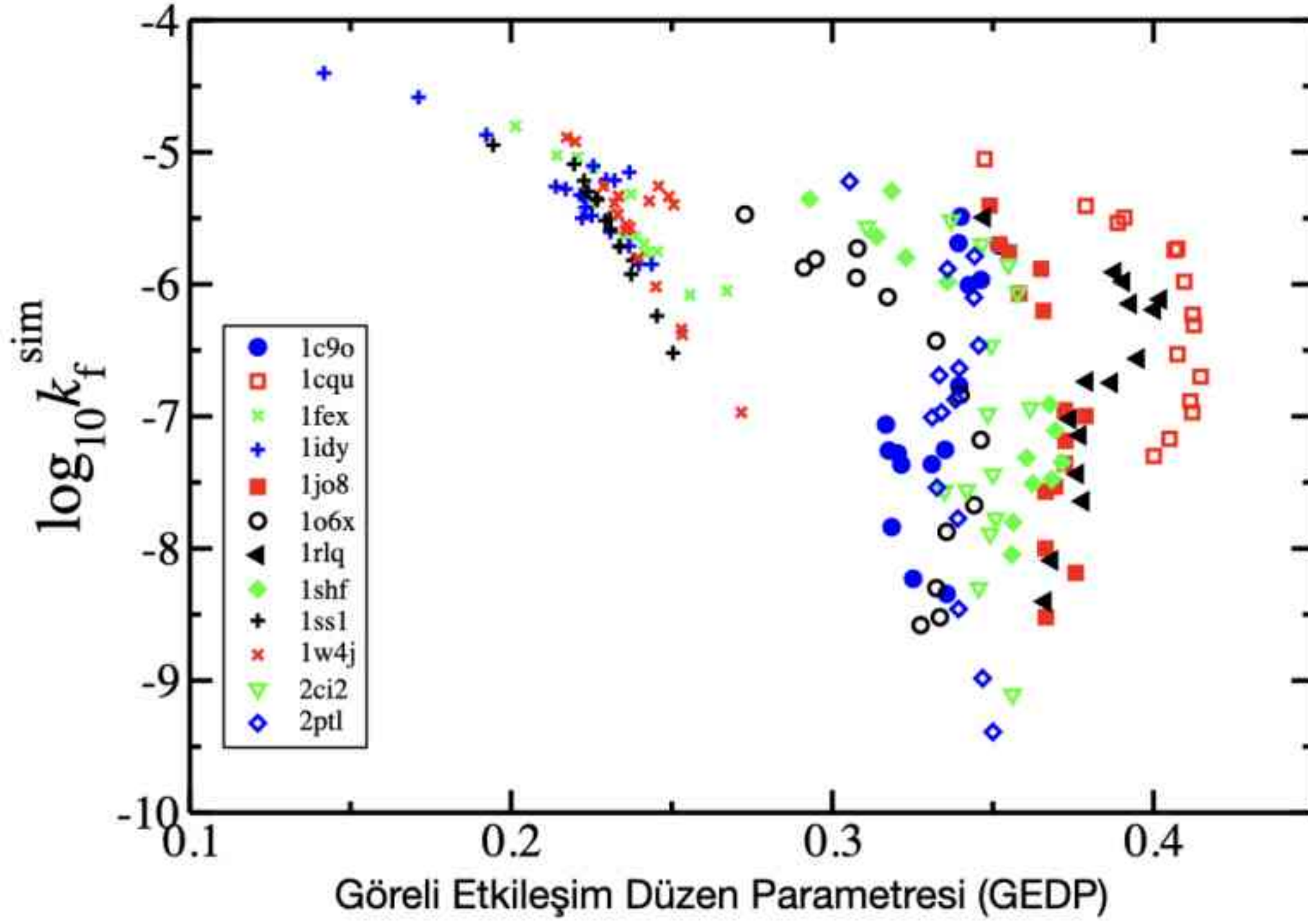
Şekil 4.5 : Simülasyon katlanma oranları ile mutlak etkileşim düzen parametresi arasındaki ilişki

MEDP parametresi proteinlerin büyüklüklerini göze önüne almaksızın bir ortalama değer hesaplamaktadır. Elde edilen sonuçları protein büyüklüğü ile ilişkilendirmek için "*görelî etkileşim düzen parametresi*" (GEDP) ayrıca tanımlanmıştır. Yani, MEDP değerleri elde edildikleri proteinlerin zincir uzunluğuna bölünerek normalize edilmektedir (bkz. Eşitlik 4.4).

$$\text{GEDP} = \frac{1}{L} \times \text{MEDP} \quad (4.4)$$

Litaratürde tanımlanmış farklı etkileşim düzen parametreleri bulunmaktadır [32, 69, 70]. Bu farklı tanımların, deneysel katlanma oranları ile olan ilişkisi Şekil 4.4'de gösterilmiştir. Şekilde de görülebileceği üzere, farklı etkileşim düzen parametreleri ile deneysel katlanma oranları arasında 0.68 ile 0.84 arasında değişen korelasyon katsayısı değerleri elde edilmiştir. Tezin bu kısmında yaygın olarak kullanılan MEDP ve GEDP etkileşim düzen parametrelerinin, aynı protein için farklı etkileşim ağırları kullanıldığında katlanma oranlarında gözlemlenen muazzam farklılığı betimleyip betimleyemeyeceği incelenecektir.

Şekil 4.5'te, bütün proteinlerin etkileşim ağları için hesaplanan MEDP değerlerine karşılık simülasyonlardan elde edilen katlanma oranları çizdirilmiştir. Termodinamik kooperativite davranışlarında gözlemlendiğimiz aynı protein ailelerine mensup proteinlerin aynı bölgelerde kümelenme davranışı, kinetik sonuçlarda da (MEDP için Şekil 4.5 ve GEDP için Şekil 4.6'ya bakınız) gözlemlenmektedir. İki şekilde gözlemlenen kümelenme kısmi farklılık gösterse de benzerlerdir.



Şekil 4.6 : Simülasyon katlanma oranları ile Görelî etkileşim düzen parametresi arasındaki ilişki

Tamamıyla α yapılı proteinlerin etkileşim ağları, çok sayıda zincir üzerinde birbirine yakın bağırsız etkileşimler (mesela, $(i : i + 4)$ vd. etkileşimler) içerdiği için MEDP ve GEDP değerleri diğer protein ailelerine mensup proteinlerden çok daha düşük çıkmaktadır. Bunun dışında düşük r_{ese} değerlerinde elde edilen etkileşim ağlarındaki toplam etkileşim sayıları diğer protein ailesine mensup proteinlerle kıyaslandığında; tamamıyla α yapılı proteinler için daha düşüktür. Bu durum protein ailelerinin farklı bölgelerde kümelenmelerinin altında yatan temel sebebi oluşturmaktadır.

MEDP ve GEDP etkileşim düzen parametreleri için varılabilecek en önemli ortak sonuçlar: (i) her iki parametre de aynı proteinin farklı etkileşim ağları için gözlemlenen katlanma oranlarındaki değişkenliği yakalamaktan çok uzaktırlar. Mesela 2ptl proteini için katlanma oranlarında $10^{4.17} \approx 14,800$ katlık farklılık gözlemlenirken, Şekil

4.5'e bakıldığında MEDP değerinde minimal bir değişkenlik olduğu (sanki dikey bir çizgi etrafında küçük dalgalanma davranışının varlığı) görülecektir. (ii) bir düzen parametresinden beklenen davranış; proteinden ve etkileşim ağlarından bağımsız olarak elde edilen tüm verilerin üst üste düşecek ve/veya birbirinin devamı/uzantısı olmasını yüksek korelasyon verecek şekilde sağlaması beklenir. Maalesef hem MEDP hem de GEDP etkileşim düzen parametreleri bu beklentileri karşılamaktan çok uzaktır.

Her ne kadar deneysel katlanma oranları ile başarılı sayılabilecek korelasyonlar verseler de; MEDP ve GEDP parametreleri sadece proteinlerin katlı yapı topolojisini betimleyebilmekte fakat aynı protein için tanımlanan farklı etkileşim ağlarının katlanma kinetiğine katkısını betimleyememektedir. Etkileşim düzen parametresi hakkında yapılan ilk çalışmada [32], aynı protein seti için aynı r_{ese} değerleri kullanılarak elde edilen etkileşim ağları için r_{ese} değeri değiştirildiğinde de benzer korelasyon gözlemlenmesinin sebebi de katlı yapı topolojisini betimlemekteki başarısıdır. Unutulmamalıdır ki; katlı yapı analizleri statik analizleridir. Bu analizlerden elde edilen sonuçların dinamik konformasyonel geçişler içeren model yaklaşımları ile test edilmesi gerekir. Yaptığımız çalışma da bunu ispatlar nitelikte olup, MEDP ve GEDP etkileşim düzen parametrelerinin düşünüldüğü kadar başarılı olmadığını göstermektedir.

4.1.3 Açısal Etkileşimli Etkileşim Düzen Parametresi

Literatürde bulunan etkileşim düzen parametre tanımlarının, protein içi etkileşim ağları için simülasyonlardan elde edilen katlanma oranlarındaki değişkenliği tanımlayamaması, bu parametrelerde ki eksikliğin nelerden kaynaklandığını incelemeye başlamanızı sağlamıştır. Etkileşim düzen parametreleri sadece $j > i + 3$ kriterini sağlayan bağımsız etkileşimleri göz önüne almaktadır. Diğer $i \leq j - 3$ kriterini sağlayan ve protein zinciri üzerinde yerel olan etkileşimleri ihmal etmektedir. Halbuki, moleküler dinamik simülasyonumuzda zincir üzerinde ardışık amino asitler arasında tanımlanan gerilme, bükülme ve burulma gibi açısal etkileşim enerji terimleri kullanılmaktadır. Her üç açısal etkileşim enerji terimi parçacıklar üzerine etkiyen kuvvetlere ve konformasyonel enerjiye (dolayısıyla tüm dinamiğe) katkı sağlamaktadır. Bu nedenle, zincir üzerinde yerel olan açısal terimlerin etkileşim düzen parametresinde ihmal edilmeyip, göz önünde bulundurulması gerekmektedir.

Bu amaçla, etkileşim ağına açısal etkileşimleri de dahil ederek yeni bir etkileşim düzen parametresi tanımladık ve bu yeni parametreyi "*Açısal Etkileşimli Etkileşim Düzen Parametresi*" (AEEDP) olarak isimlendirdik. Eşitlik 4.5’de matematiksel formu verilen AEEDP parametresi, protein dinamiğine katkı sağlayan tüm açısal terimlerdeki (yani, eğilme, gerilme ve bükülme enerji terimleri) ve katlı yapıda bağımsız etkileşim içerisinde bulunan tüm amino asit çiftleri, etkileşim ağına dahil edilerek hesaplanmaktadır.

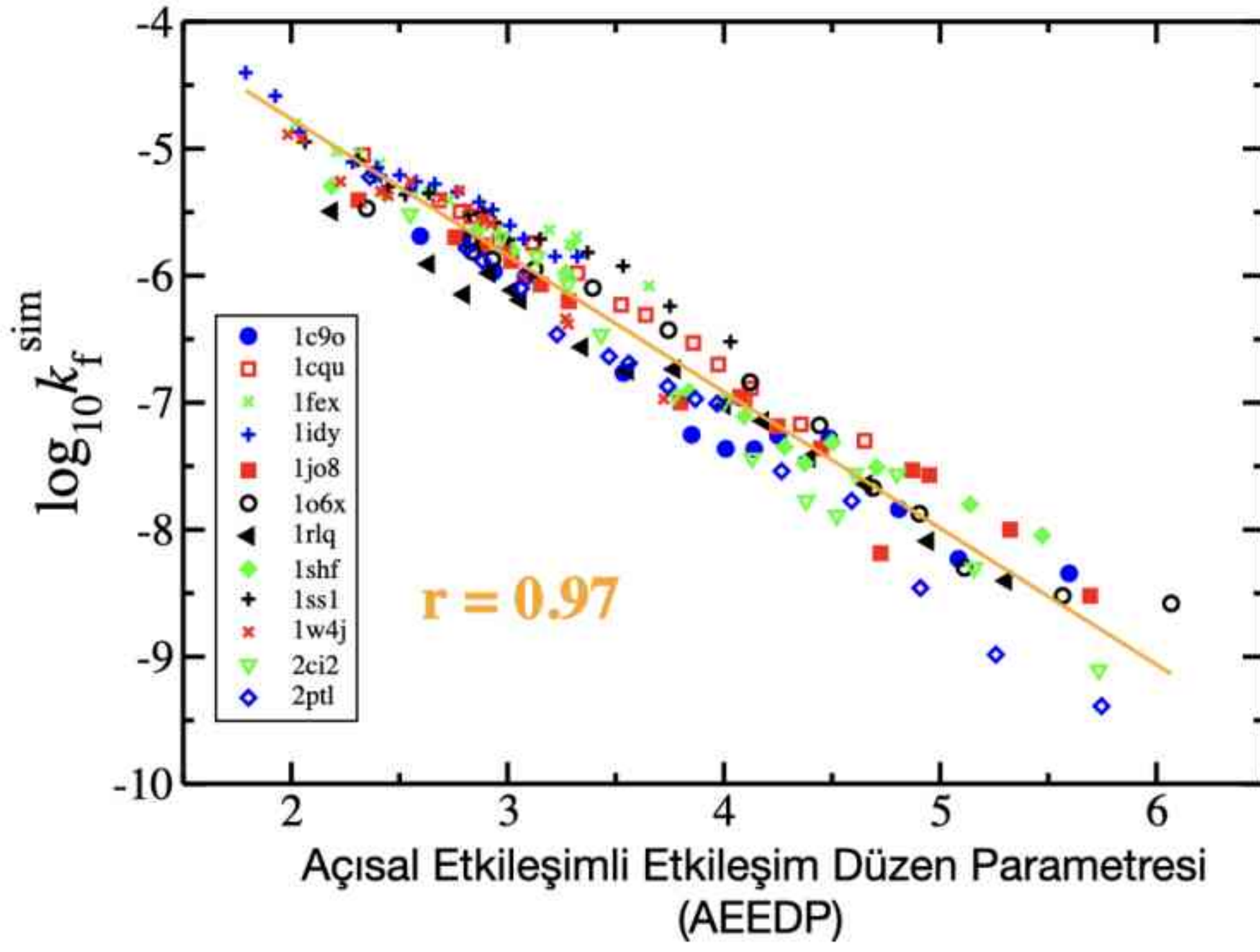
$$\begin{aligned}
 AEEDP &= \frac{\sum_{N=1}^{N-1} \Delta Z_{\text{gerilme}} + \sum_{N=2}^{N-2} \Delta Z_{\text{bükülme}} + \sum_{N=3}^{N-3} \Delta Z_{\text{burulma}} + \sum_{N_{\text{bağımsız}}^{\text{katlı}}} \Delta Z_{ij}}{N_{\text{gerilme}} + N_{\text{bükülme}} + N_{\text{burulma}} + N_{\text{bağımsız}}^{\text{katlı}}} \\
 &= \frac{25L - 69 + \sum_{N_{\text{bağımsız}}^{\text{katlı}}} \Delta Z_{ij}}{16L - 43 + N_{\text{bağımsız}}^{\text{katlı}}} \quad (4.5)
 \end{aligned}$$

Açısal etkileşimlerin protein içi etkileşim ağına eklenmesi bir önceki kısımda anlatıldığı gibi; çoklu parçacık etkileşmesi içeren büküle ve burulma enerji terimlerinden gelecek katkı çoklu parçacık etkileşmeli bir durumundan, birden fazla iki parçacık etkileşmeli duruma transformasyonu gerçekleştirilerek yapılmaktadır. Böylelikle açısal terimlerden gelen hem etkileşim sayısı hem de zincir üzerindeki mesafe değerleri kolaylıkla hesaplanabilmektedir. Eşitlik 4.5’de görüleceği üzere açısal terimlerin etkileşim ağındaki toplam etkileşme sayısına sağlayacağı katkı $16L - 43$ ve etkileşen amino asit çiftlerinin zincir üzerindeki toplam mesafesine sağlayacağı katkı $25L - 69$ değerlerine eşit olup sadece protein zincir uzunluğuna bağlıdır. Bu katkı aynı proteinin farklı etkileşim ağlarında değişmemektedir. Bu terimler için iki parçacık tasvirinin nasıl uygulandığının detayları Şekil 2.2’de kırmızı kesikli çizgiler ile gösterilmiştir.

4.1.3.1 AEEDP etkileşim düzen parameteresi ve katlanma oranları arasındaki ilişki

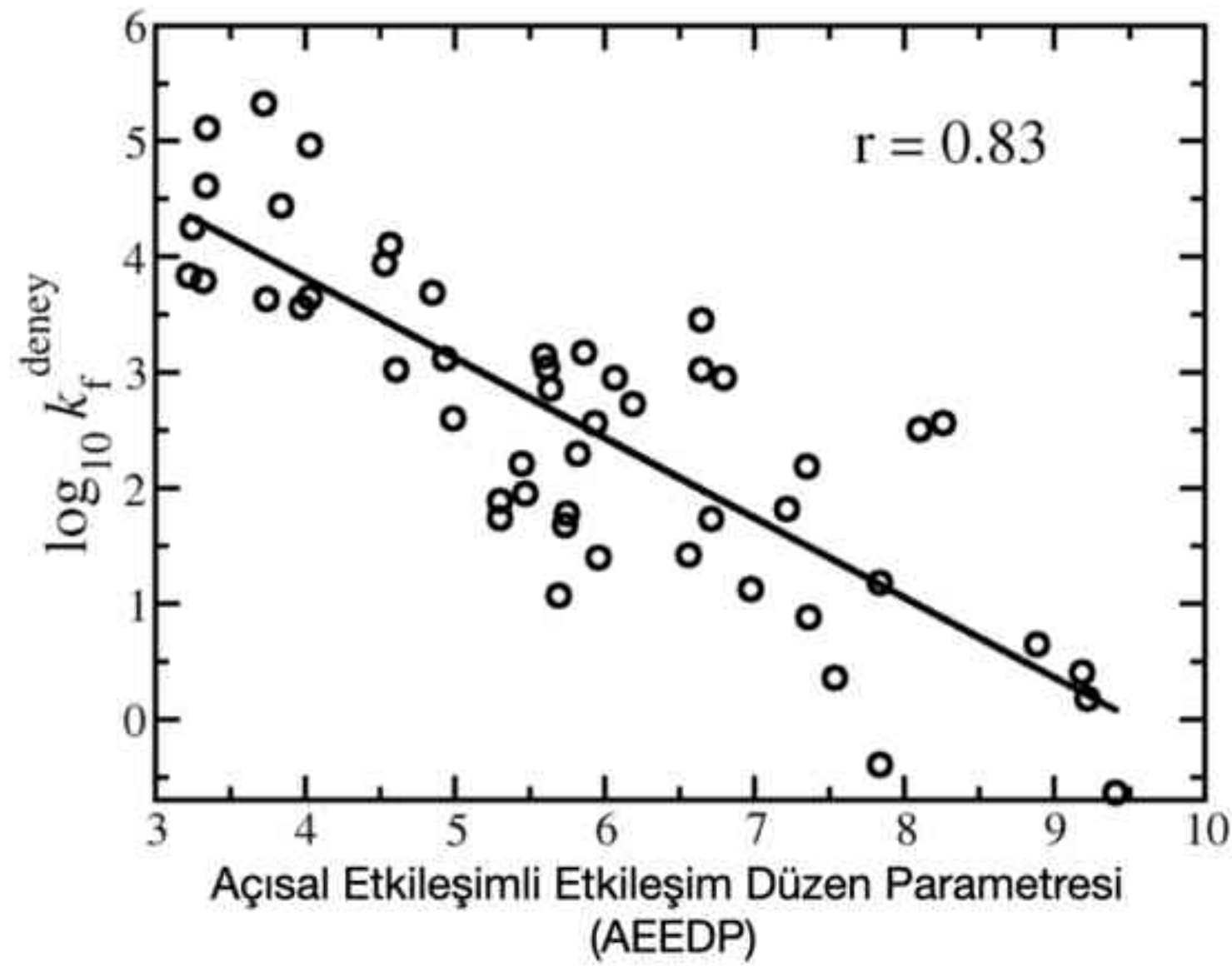
Farklı etkileşim ağları için elde ettiğimiz katlanma oranları ile yeni etkileşim düzen parametremiz, AEEDP, parametresinin ilişkisi (bkz. Şekil 4.7) incelendiğinde; MEDP ve GEDP etkileşim düzen parametrelerinden tamamen farklı davranışlar gözlemlenmektedir.

Bütün protein yapıları ve farklı protein içi etkileşim ağları için elde edilen katlanma oranlarının hepsi birbirleri üzerine düşmekte ve doğrusal bir çizgi üzerinde lineer olarak değiştiğini görmekteyiz. Diğer bir deyişle, AEEDP parametresi hem aynı proteinin farklı protein içi etkileşim ağlarından elde edilen katlanma oranı değişimini yakalamakta hem de



Şekil 4.7 : AEEDP parametresinin simülasyondan elde edilen katlanma oranları ile ilişkisi. Turuncu doğru iki değerin lineer korelasyonu göstermekte göstermekte ve korelasyon katsayısı 0.97 olarak bulunmuştur.

farklı ailelere mensup proteinlerin farklı bölgelerde kümelenmeleri yerine bir araya getirip, üst üste örtüşmelerini sağlamaktadır. AEEDP ile katlanma oranları arasında hesaplanan Pearson korelasyon [71] katsayısı 0.97 gibi mükemmel sayılabilecek bir değere sahiptir.



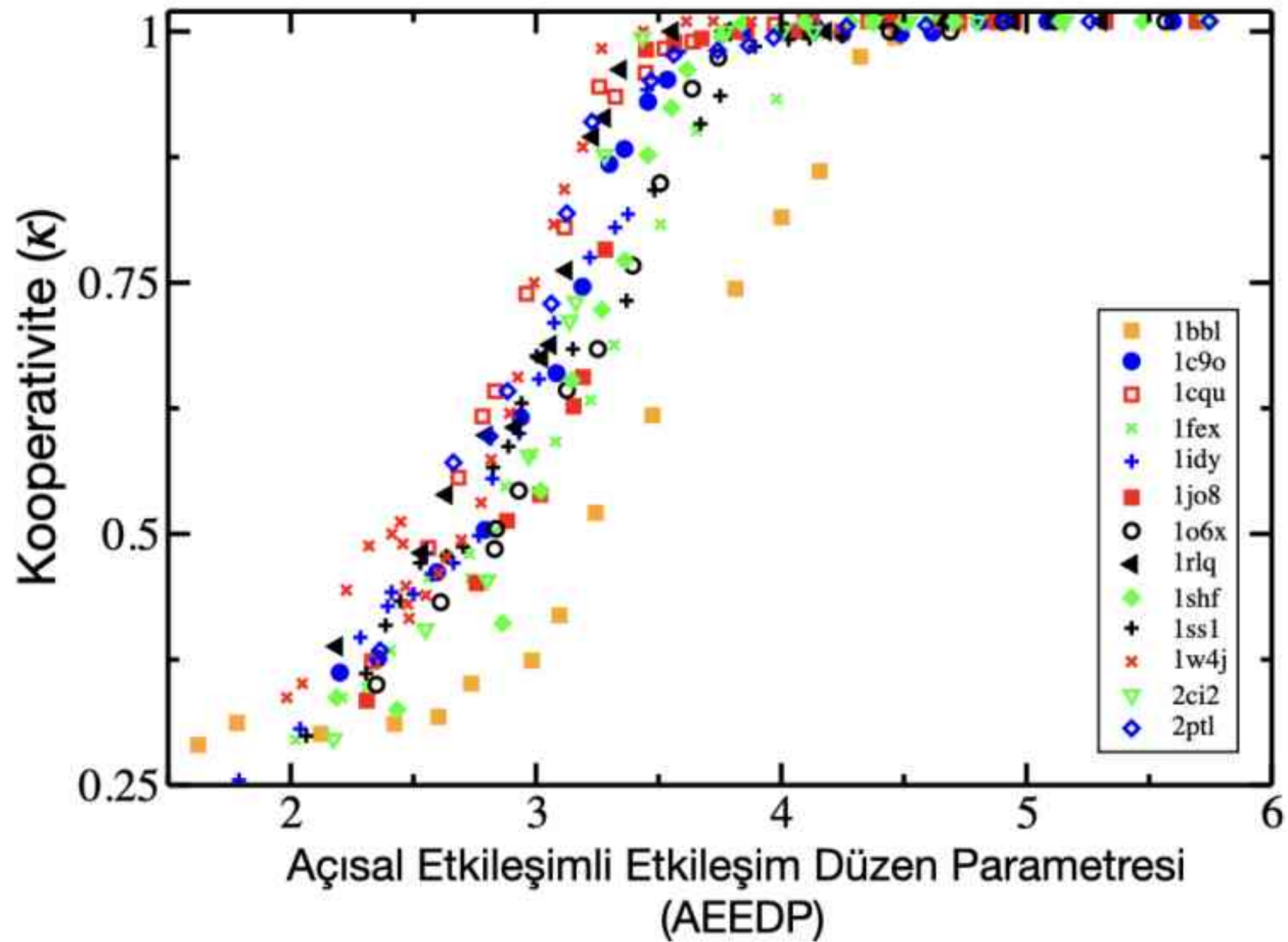
Şekil 4.8 : AEEDP parametresinin deneysel katlanma oranları ile ilişkisi.

AEEDP parametresinin simülasyon katlanma oranlarını tahmin etmekteki başarısı gösterilmiştir. Şekil 4.8’de ise AEEDP’nin deneysel katlanma oranları ile olan ilişkisi araştırılmıştır. Burada kullanılan protein seti, diğer EDP parametreleri ile birebir kıyaslanabilir olması için Şekil 4.4’te kullanılan protein seti kullanılmıştır.

Şekil 4.8’de görüldüğü üzere, geliştirdiğimiz AEEDP etkileşim düzen parametresi, deneysel katlanma oranları ile diğer EDP parametrelerinin çoğunluğundan daha yüksek korelasyon vermektedir. AEEDP için hesaplanan lineer Pearson korelasyon katsayısı 0.83’dir. Yani AEEDP hem deneysel katlanma oranları ile hem de simülasyondan elde edilen katlanma oranları ile oldukça yüksek korelasyonlar vermektedir.

4.1.3.2 AEEDP etkileşim düzen parametresi ve termodinamik kooperativite ilişkisi

Geliştirdiğimiz AEEDP etkileşim düzen parametresinin simülasyon ve deneysel katlanma oranları ile olan başarılı ilişkisi sonucunda, aynı parametrenin termodinamik kooperativite ile olan ilişkisi de ayrıca incelenmiştir.



Şekil 4.9 : Geliştirdiğimiz AEEDP etkileşim düzen parametresi ile termodinamik kooperativite arasındaki ilişkinin gösterimi.

Şekil 4.9’da, AEEDP parametresine karşı termodinamik kooperativite parametresi κ çizdirilmiştir. Daha önce incelediğimiz etkileşim sona erim mesafesi r_{ese} (bkz. Şekil 4.2) ve ortalama etkileşim sayısı OES (bkz. Şekil 4.3) parametrelerinin termodinamik

kooperativite ile olan ilişkisinde gözlemlenen protein ailesi kümelenme davranışlarının aksine, AEEDP parametresi farklı aileye mensup proteinleri bir bant aralığında toplayarak, bütün proteinlerin aynı AEEDP değerinde kooperatif davranış göstermeye başladığını işaret etmektedir. Yani protein ailesi ayrımı gözetmeksizin proteinlerin iki-durum termodinamik davranışı gösterebilmeleri için AEEDP parametresinin bir eşik değerine sahip olduğu ve bu değer minimum 3.25 olması gerektiği bulunmuştur.

Diğer parametrelerde, olduğu gibi deneysel olarak bariyersiz davranış gösteren 1bbl proteini AEEDP parametresinde de iki durum davranışı gösteren proteinlerden ayrıştığı gözlemlenmiştir.

Sonuç olarak, bu tez çalışmasında basitleştirilmiş protein zincir temsili ve Langevin dinamiği içeren moleküler dinamik simülasyon yönteminin kullanıldığı Gō modelimizden elde edilen sonuçlar; geliştirdiğimiz AEEDP etkileşim düzen parametresinin, proteinlerin hangi protein ailesine mensup olduklarından, zincir uzunluklarının ne kadar olduğundan, etkileşim ağının hangi proteine ait olduğundan bağımsız olarak deneysel sonuçlarla uyumlu olarak proteinlerin hem termodinamik hem de kinetik davranışlarını eş zamanlı olarak betimleyebildiği bulunmuştur.

4.2 Tüm Atomların Temsil Edildiği Monte Carlo Protein Modeli

Tezin bu kısmına kadar katlı yapısı deneylerle elde edilmiş proteinler kullanılarak protein termodinamiği ve kinetiği incelenmiştir. Bir önceki bölümde elde edilen sonuçlardan anlaşılabileceği üzere; basitleştirilmiş zinciri temsilli protein spesifik model yaklaşımları kullanarak protein katlanması ve kinetiği davranışları hakkında çok şey öğrenebiliyor ve davranış farklılıklarını betimleyebilecek parametreler ve çözümler üretebiliyoruz. Ama protein katlanması alanında ulaşılmak istenen en büyük hedef: proteinlerin tüm atomlarının temsil edildiği, transfer edilebilir (yani, protein spesifik olmayıp tüm proteinler için kullanılabilecek) etkileşim potansiyelleri içeren *ab-initio* modeller geliştirmektir. Böylelikle katlı yapıları bilinmese dahi, proteinlerin sadece amino asit dizilimi bilgisi kullanılarak; protein katlı yapılarının tahmin edilmesi, katlı yapıya ulaşırken gerçekleştirdikleri dinamik konformasyonel geçişlerin takip edilebilmesi ve dolayısıyla protein dizi-yapı-kararlılık-fonksiyon ilişkilerinin ortaya çıkarılması mümkün olacaktır. Proteinlerin yanlış katlanması ve/veya yanlış bir araya gelmesi sonucu oluşan hastalıkların [72] sayısındaki büyük artış; bu doğrultuda yapılacak bilimsel çalışmaların

ve elde edilebilecek başarılı sonuçların yaratabileceği yaygın etkinin büyüklüğünü ve önemini işaret etmektedir.

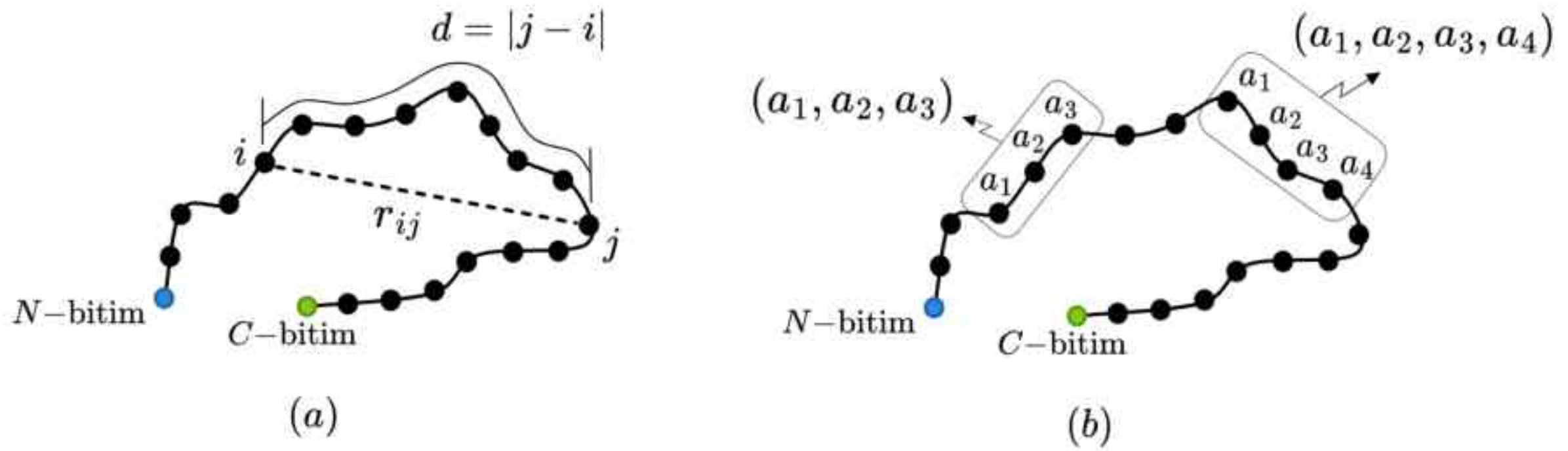
NMR [21], X-ışını [20] kristalografi, elektron mikroskobu [22] ve diğer deneysel yöntemler ile şu ana kadar 200,000 civarında proteinin 3-boyutlu katlı yapısı çözümlenmiş olsa dahi; canlı organizmalarda bulunan ve amino asit dizilimleri (dizi bilgisi) bilinip fonksiyonel olarak aktif oldukları katlı yapıları bilinmeyen milyarlarca protein dizisi mevcuttur. Ağustos 2021’de Nature dergisinde yayınlanan bilimsel çalışmada [73]; Protein Veri Bankası (PVB)’nda mevcut olan deneysel çalışmalardan elde edilmiş 170,000 proteinin katlı yapılarından elde edilen evrimsel, fiziksel ve geometrik sınırlayıcı bilgilerin, eğitim verileri olarak kullanıldığı AlphaFold adında bir Yapay Zeka algoritması oluşturulmuştur. Bu yapay zeka algoritması ile amino asit dizilimi verilen bir proteinin katlı yapısı %94 doğruluk oranı ile tahmin edilebilmektedir. Protein katlı yapı tahminindeki bu gelişme gerek ilaç dizaynı ve gerekse protein fonksiyonlarının anlaşılmasında çok kritik rol oynamaktadır. Bununla birlikte, AlphaFold vb. yapay zeka algoritmaları [73–76], proteinlerin dinamik davranışları ve katlı yapıya ulaşmakta hangi fiziksel mekanizmaların etkin olduğu hakkında maalesef herhangi bir bilgi sunmamakta, diğer bütün yapay zeka algoritmalarında olduğu gibi kara kutu (black box) şeklinde çalışmakta, ve tek bir amaca yönelik (sadece katlı yapı tahmini vb) tahmin yapmaktadır.

Diğer taraftan, şu ana kadar geliştirilen bilgisayara dayalı protein modelleri ve kuvvet alanları, proteinlerin katlı yapılarına ulaşılırken geçirdikleri dinamik konformasyonel değişimleri tasvir etmekten, bu dinamik dönüşümlerde hangi amino asitlerin kritik rol üstlendiğini belirlemekten ve daha da önemlisi deneysel termodinamik ve kinetik davranışları ihtiva etmekten hala uzaktır. Bir protein için başarılı sonuçlar veren bir model, başka bir protein için başarısız sonuçlar vermektedir. Bu nedenle, bir taraftan literatürde kabul görmüş AMBER [10], CHARMM [8, 77] etkileşim potansiyelleri ve kuvvet alanlarının yeni versiyonlarının geliştirilmesine devam edilirken, diğer taraftan yeni enerji terimlerinin önerilmesine ve test edilmesine devam edilmektedir.

4.2.1 Transfer Edilebilir İstatistiksel Etkileşim Potansiyel Dizaynları

Protein veri bankasındaki mevcut protein yapılarının; kullanılan deneysel yöntem, çözünürlük, zincir özdeşliği, R-faktörü ve protein zincir uzunluğuna göre seçip sınıflandıran PISCES [78] sunucusundan ([http:](http://)

//dunbrack.fccc.edu/lab/pisces) elde edilen protein setlerinden cullpdb-pc50.0-res0.0-2.5-len40-10000-R0.3-Xray-d2022-01-26-chains25982.txt dosyasında belirtilen 25,982 protein ele alınmıştır. Bu protein setini oluşturan proteinlerin hepsi X-ışını kristalografi deneysel yöntemi kullanılarak elde edilmiş, zincir özdeşlikleri maksimum %50, kristal yapı çözünürlüğü 2.5 Å veya daha düşük, protein zincir uzunluğu en az 40 amino asit olan, R-faktörü 0.3 olan proteinlerden oluşturulmuştur. Protein yapı çözünürlüğünü özellikle düşük seçmemizin sebebi kristal yapı hatalarını düşük tutabilmektir. Buna rağmen protein seti üzerinde yaptığımız detaylı incelemede; standart olmayan amino asitlere sahip 171 protein analizimizin dışında bırakılmıştır. Diğer bir deyişle, protein veri setimiz 25,801 proteinden oluşmakta olup, toplam amino asit sayısı 6,473,730 ve toplam atom sayısı 50,607,497'dir.



Şekil 4.10 : Protein zinciri üzerinde; (a) i 'nci ve j 'inci amino asitler arası öklid mesafesi r ve protein zinciri üzerindeki mesafe tanımının $d = |i - j|$, ve (b) ardışık amino asit üçlü ve dörtlülerinin notasyonları ve zincir üzerinde gösterimi.

Bu tez kapsamında geliştirilen transfer edilebilir etkileşim potansiyellerinin anlatımını kolaylaştırmak için yaptığımız protein bileşen tanımları ve kullandığımız değişken ve parametre temsilleri ve ilgili notasyonlar şu şekilde açıklanabilir: $\tilde{a}$ ve a terimleri sırasıyla atom ve amino asit türlerini temsil etmektedir. Amino asit üçlü ve dörtlüleri sırasıyla (a_1, a_2, a_3) ve (a_1, a_2, a_3, a_4) notasyonları ile gösterilmekte ve alt indisler Şekil.4.10.(b)'de gösterildiği gibi amino asitlerin üçlü veya dörtlü içindeki sırasını vermektedir. Bu sıralama amino asitlerin, proteinin N -bitimine yakınlığına göre düzenlenmektedir. Yani, N -bitime en yakın olandan, en uzak olana doğru sıralama yapılmaktadır. Benzer tanımlama atom çifti $(\tilde{a}_1, \tilde{a}_2)$ 'leri içinde geçerlidir. Verilen bir $\tilde{a}$ atomunun hangi amino asite ait olduğunu belirtmek için a sol alt indisi (yani, $a\tilde{a}$) kullanılacaktır. Spesifik bir örnek vermek gerekirse; mesela alanin amino asitinin C_α atomu, AC_α şeklinde gösterilecektir. İlgili atom, ikili, üçlü ve dörtlülerdeki sırası ile gösterilmek istendiğinde mesela $AC_{\alpha 1}$ notasyonu

kullanılacaktır. Diğer taraftan, atom çiftleri için uzaysal mesafe r ; iki atomun türü ile değil atomların, protein zincirinde N -bitimine yakınlığına göre düzenlenmiş atom çifti içindeki sırasına göre; yani, $(\tilde{a}_1, \tilde{a}_2)$ çiftinde $\tilde{a}_1$ ve $\tilde{a}_2$ atomları arasındaki mesafe r_{12} olarak ifade edilecektir. Ayrıca, ölçülen uzaysal r_{12} mesafeleri 0.1 Å genişliğinde aralıklara ayrılarak, süreksiz/ayrık bir değişken haline dönüştürülmüştür.

4.2.1.1 Atom çiftleri için transfer edilebilir istatistiksel etkileşim enerji terimi dizaynı

Proteinleri oluşturan 20 farklı amino asitin her atomu farklı bir atom çeşidi olarak ele alınmış, böylelikle toplam 167 farklı atom türü tanımlanmıştır. Amino asitlerin omurga atomlarının simetrik olmayan doğası dolayısıyla; atomlar arasında mesafeler ve oluşturdıkları geometrilerde asimetric davranışlar gözlemlenebilmektedir. Mesela, ${}_AN$ ve ${}_AO$ atomları dikkate alındığında; bunların oluşturabileceği atom çiftlerinin uzaysal mesafe dağılımları birbirinden farklıdır, matematiksel gösterimle $({}_AN_1, {}_AO_2) \neq ({}_AO_1, {}_AN_2)$ 'dir. Bu nedenle, toplam atom çifti çeşidi sayısı $167 \times 167 = 27,889$ 'dur. Belirtilmesi gerekir ki, zincir üzerinde birbirine çok yakın (yani, yerel) olan amino asitlere ait atomların gerek sterik şartları yerine getirebilecekleri gerekse etkileşim enerjilerini minimize edebilecekleri olası konfigürasyon sayısı oldukça düşüktür. Atom çiftlerini oluşturan atomların zincir üzerindeki uzaklıkları arttıkça sahip olacakları serbestlik dereceleri de artmaktadır. Protein katlı yapıları incelendiğinde, zincir üzerinde yerel olan amino asit çiftlerinin omurga atomları; sarmallar (heliks) ve β -plakaları gibi belirli geometrilere sahip ikincil yapılar oluşturmaktadırlar. Amino asitlerin birbirlerine zincir üzerinde olan mesafeleri $d = |i - j|$ parametresi (bkz. Şekil.4.10.(a)) ile tanımlanırsa ve ardından atom çiftlerindeki atomların ait oldukları amino asitlerin d mesafelerine göre ayrı ayrı sınıflandırıldığında; atom çiftlerinin öklid mesafesi r 'nin bir fonksiyonu olarak gözlemlenme sıklık/frekans dağılımlarının çok yüksek değişkenlikler gösterdiği bulunmuştur. Daha detaylı açıklamak gerekirse, protein zinciri üzerinde yerel olarak tanımladığımız $d = 1, 2, 3, 4, 5, 6$ mesafe değerlerinin herbirinde atom çiftleri için birbirinden oldukça farklı gözlemlenme frekans dağılımları, $d > 6$ olduğu durumlarda ise benzer dağılımlar elde edilmiştir. Bu nedenle, $d > 6$ olan tüm durumlar aynı veri setinde tek bir durum olarak ele alınmıştır. Sonuç olarak, tüm atom çiftleri $d = 1, 2, 3, 4, 5, 6$ ve $d > 6$ olacak şekilde 7 farklı zincir mesafesine göre ayrı ayrı sınıflandırılarak ele

alınmıştır. Böylelikle, tanımlanan farklı atom çifti çeşidi veya etkileşme türü sayısı $7 \times 167 \times 167 = 195,223$ 'e çıkmıştır.

Dengede olan termodinamik sistemler prensip olarak serbest enerjilerini minimum yapmaya çalışırlar. Bu nedenle, denge durumunda proteinlerin katlı yapılarının minimum serbest enerjiye sahip olduğu ve her bir mikroskobik duruma karşılık gelen protein konformasyonlarının dağılımı Boltzmann prensibine uyduğu kabul edilmektedir [79]. Boltzmann kanunu bir sistemin enerjisi, o sistemin olasılık yoğunluk fonksiyonu ile ilişkilendirmekte ve süreksiz/ayrık değişkenler kullanıldığında; olasılık yoğunluk fonksiyonu Eşitlik 4.6 ile ifade edilmektedir.

$$P_{ijk} = Z^{-1} \exp\left(\frac{-E_{ijk}}{k_B T}\right) \quad (4.6)$$

Burada k_B Boltzmann sabiti, T sıcaklık ve i, j, k alt indisleri sistem değişkenlerini ifade etmektedir. Z terimi ise Boltzmann toplamı veya bölüşüm fonksiyonu olarak adlandırılmakta ve matematiksel formu Eşitlik 4.7 ile verilmektedir.

$$Z = \sum_{ijk} \exp\left(\frac{-E_{ijk}}{k_B T}\right) \quad (4.7)$$

Bu çalışmada Boltzmann kanunu ve protein katlı yapı koordinatları kullanılarak; atom çiftleri için $P_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12})$ olasılık yoğunluğu tanımlanmış ve aşağıda eşitlikteki gibi hesaplanmıştır.

$$P_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12}) = \frac{\frac{N_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12})}{\sum_n N_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12}^n)}}{\frac{\sum_{i=1}^{167} \sum_{j=1}^{167} N_{\text{çift}}^d(\tilde{a}_1^i, \tilde{a}_2^j, r_{12})}{\sum_{i=1}^{167} \sum_{j=1}^{167} \sum_n N_{\text{çift}}^d(\tilde{a}_1^i, \tilde{a}_2^j, r_{12}^n)}} \quad (4.8)$$

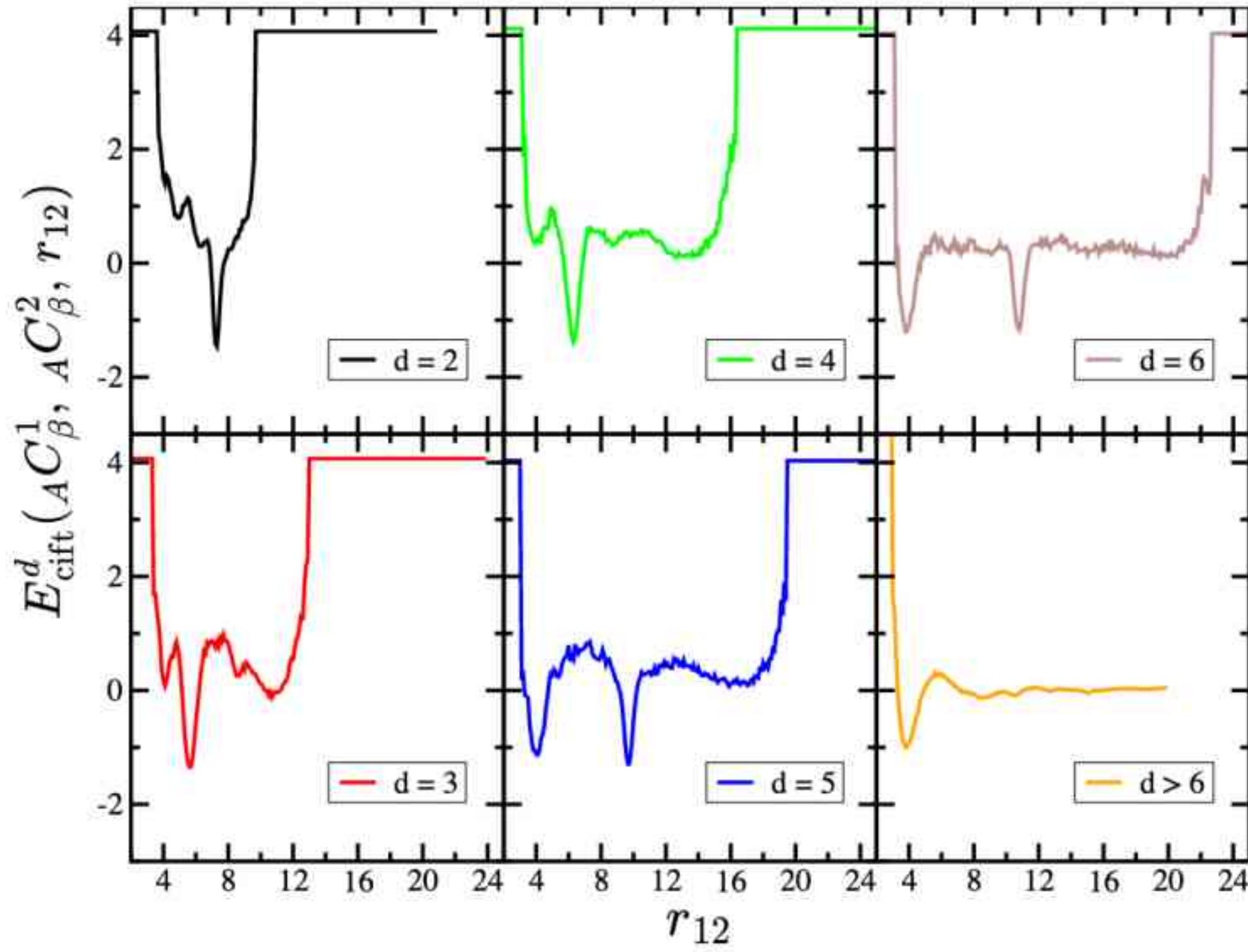
Eşitlik.4.8'de d parametresi $(\tilde{a}_1, \tilde{a}_2)$ atom çiftindeki $\tilde{a}_1$ ve $\tilde{a}_2$ atomlarının ait oldukları amino asitlerin zincir üzerindeki mesafelerini belirten değişkeni (bkz. Şekil.4.10.(a)), $N_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12})$ terimi ise d zincir mesafesinde ve r_{12} öklid mesafesinde gözlemlenmiş toplam $(\tilde{a}_1, \tilde{a}_2)$ atom çifti sayısını vermektedir. Yine, eşitlikteki i ve j indisleri farklı tüm atom türlerinin (167 tane) belirteci olup, i ve j üzerinden yapılan iki toplam tüm atom çiftlerinin hesaba katıldığını göstermektedir. n üst indisi ise yine d zincir mesafesindeki $(\tilde{a}_1, \tilde{a}_2)$ atom çifti için gözlemlenen ayrık öklid uzayı mesafesi r_{12} 'nin ait

olduğu aralık veya bin belirteci olup, n üzerinden yapılan toplam gözlemlenmiş tüm r_{12} ayırık mesafelerini kapsadığını göstermektedir.

Bu tanımlamalar altında her $(\tilde{a}_1, \tilde{a}_2)$ atom çiftinin d ve r_{12} mesafelerindeki enerjisi $E_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12})$ aşağıdaki Eşitlik 4.9'a göre hesaplanmıştır [80].

$$E_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12}) = k_B T \ln \left(1 + \sigma m_{\tilde{a}_1 \tilde{a}_2}^d \right) - k_B T \ln \left(1 + \sigma m_{\tilde{a}_1 \tilde{a}_2}^d P_{\text{çift}}^d(\tilde{a}_1, \tilde{a}_2, r_{12}) \right) \quad (4.9)$$

Burada σ sabiti gözlemlenen her bir atom çiftinin istatistiksel ağırlığı olup, değeri $1/50$ olarak alınmıştır. $m_{\tilde{a}_1 \tilde{a}_2}^d$ parametresi ise d zincir mesafesinde gözlemlenen $(\tilde{a}_1, \tilde{a}_2)$ atom çiftinin toplam sayısını temsil etmektedir.



Şekil 4.11 : Alanin amino asitinin β -karbon (AC_β) atomları arasında farklı d zincir mesafelerinde elde edilen etkileşim enerjilerinin öklid mesafesi r_{12} 'ye göre değişimi.

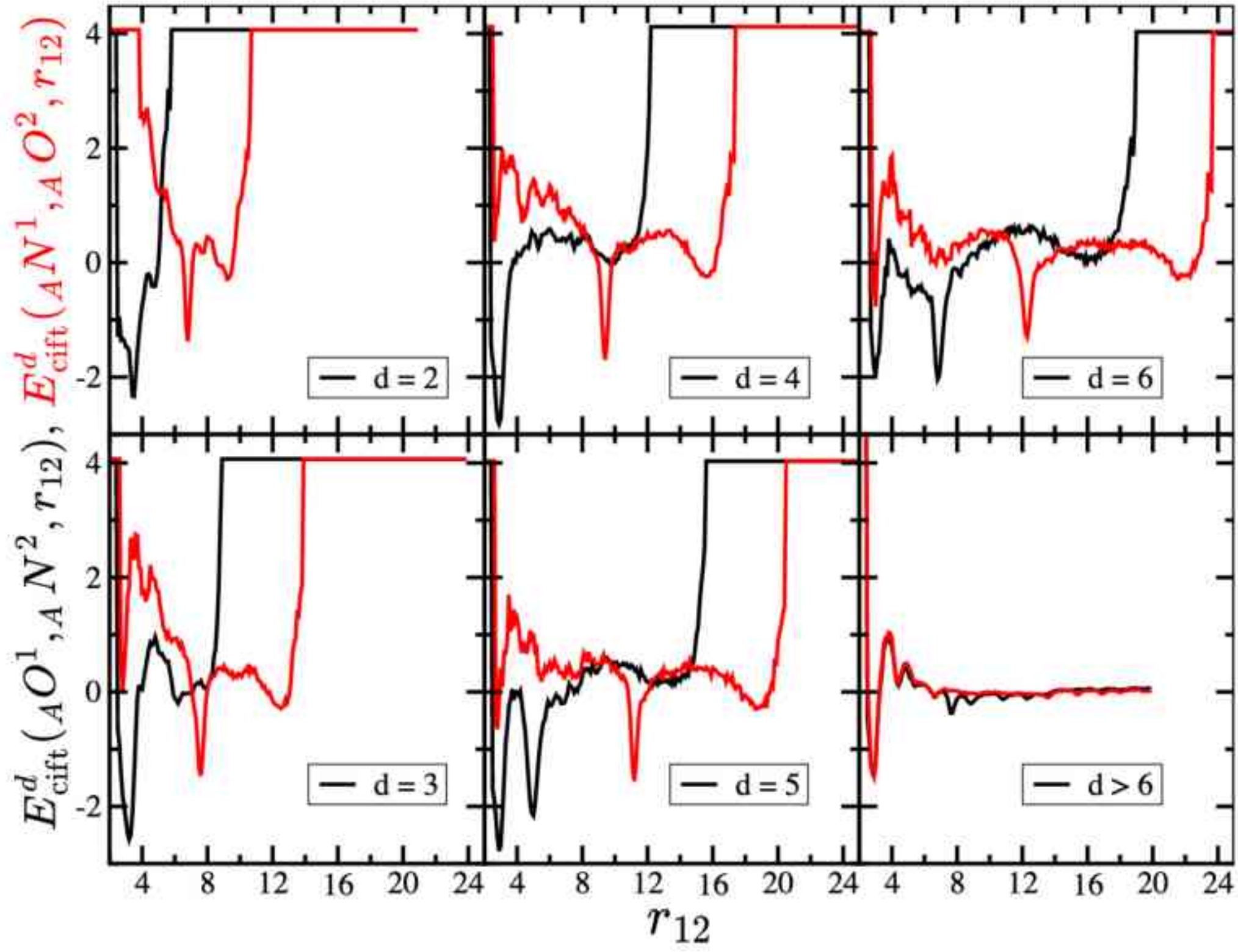
Şekil.4.11'de görüleceği üzere alanin amino asitinin β -karbon atomu ile başka bir alaninin β -karbon atomunun oluşturduğu (AC_β, AC_β) atom çiftinin farklı d zincir mesafelerinde hesaplanan etkileşim potansiyellerinin öklid mesafesine göre değişimleri incelendiğinde; çok şaşırtıcı sonuçlar elde edilmiştir. Her d değerinde elde edilen minimum etkileşim enerji değeri ve bu değere karşılık gelen r_{12} değerleri yüksek değişkenlik göstermektedir. Ayrıca $d = 2, 3, 4$ zincir mesafelerinde tek bir global enerji minimumu ve bir potansiyel

kuyusu mevcut iken $d = 5, 6$ zincir mesafelerinde ise iki global enerji minimum ve iki potansiyel kuyusunun varlığı gözlemlenmektedir.

Bu davranış farklılıklarını belirleyen en temel sebepler ise: (i) dengedeki sistemlerin enerjilerini minimum yapacak konformasyon tercihleri, (ii) atom çiftlerinin protein zinciri üzerindeki yakınlık derecesi, (iii) atomların birbirine kovalent bağ ile bağlı olmasının beraberinde getirdiği geometrik kısıtlılık derecesidir. Protein zinciri üzerinde yakınlık derecesi azaldıkça serbest atom çifti davranışı gözlemlenmektedir. Mesela, yerel olmayan ve geometrik kısıtlılığın minimal olduğu $d > 6$ zincir mesafeleri için elde edilen sonuçlar; yarı-ampirik (*semi-empirical*) yöntemlerle elde edilen sonuçlarla uyumludur. Diğer bir deyişle, atom çiftlerinin van der Waals yarıçaplarının toplamı denge mesafesine karşılık gelmekte ve bu denge mesafesinde atom çiftinin etkileşim enerjisi minimum değerine sahip olmaktadır. Şekil.4.11'de ele alınan $({}_AC_\beta, {}_AC_\beta)$ atom çiftinin van der Waals yarıçaplarının toplamı $(1.88 + 1.88 = 3.76\text{\AA})$ olup, hesaplamalarımızın $d > 6$ için verdiği değer ise $3.8\text{\AA}$ 'dur. Ayrıca, iki atom arasındaki mesafe, denge mesafesine yaklaşırken aralarında bulunan su molekülünü dışarlayabilmek için aşılması gereken su dışarlama bariyeri (desolvation barrier) [40, 81] varlığının gözlemlenmesi yaklaşımlarımızın ve hesaplarımızın ne kadar başarılı olduğunu göstermektedir. Görüleceği üzere, aynı atom çifti $({}_AC_\beta, {}_AC_\beta)$ farklı d zincir mesafelerinde farklı fiziko-kimyasal özellikleri olan farklı atom çiftleri gibi davranmaktadırlar.

4.2.2 Yönelim ve düzlemsellik özelliği olan transfer edilebilir istatistiksel hidrojen bağı enerji terimi dizaynı

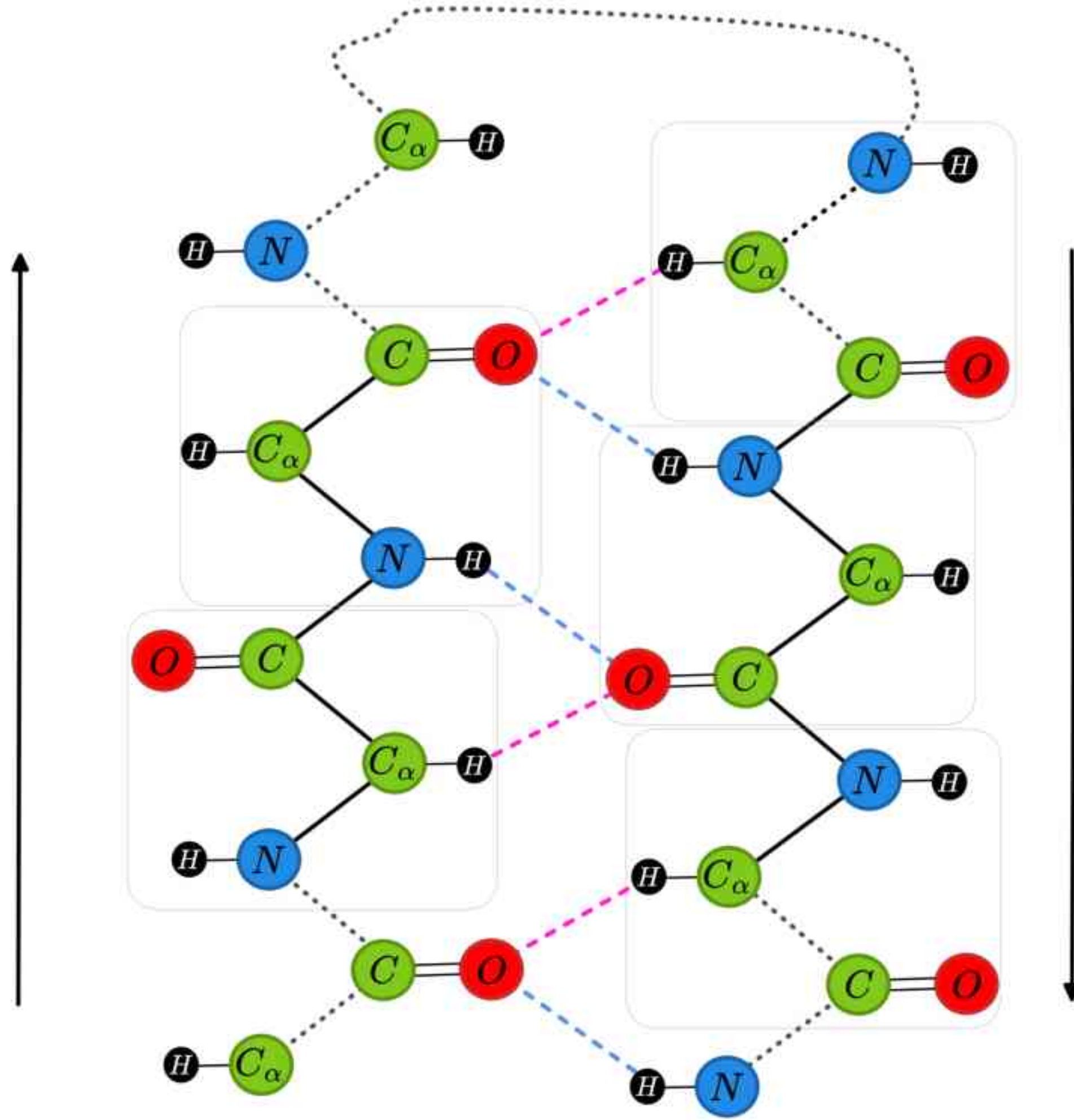
Bilindiği üzere oksijen ve azot atomlarının sahip oldukları yüksek elektronegatiflik dolayısıyla azot atomuna bağlı donör hidrojen atomu ile akseptör oksijen atomu arasında kolaylıkla hidrojen bağı oluşabilmektedir [82]. Bu hidrojen bağları hem ikincil yapı oluşumuna hem de protein yapı kararlılığına büyük katkı sağlamaktadır [83]. Alanin amino asitinin $({}_AO_1, {}_AN_2)$ ve $({}_AN_1, {}_AO_2)$ atom çiftleri için aynı d zincir mesafe değerlerinde elde edilen $E_{\text{çift}}^d({}_AO_1, {}_AN_2, r_{12})$ ve $E_{\text{çift}}^d({}_AN_1, {}_AO_2, r_{12})$ enerji değerlerinin r_{12} öklid mesafesine göre değişimi incelendiğinde (bkz. Şekil.4.12'de); siyah renkli çizgi ile gösterilen ve ${}_AO$ atomunun zincirin N-bitimine daha yakın olduğu $({}_AO_1, {}_AN_2)$ atom çifti ile kırmızı renkli çizgi ile gösterilen ve ${}_AN$ atomunun zincirin N-bitimine daha yakın olduğu $({}_AN_1, {}_AO_2)$ atom çiftinin aynı $d = 2, 3, 4, 5, 6$ zincir mesafelerinde ne kadar yüksek



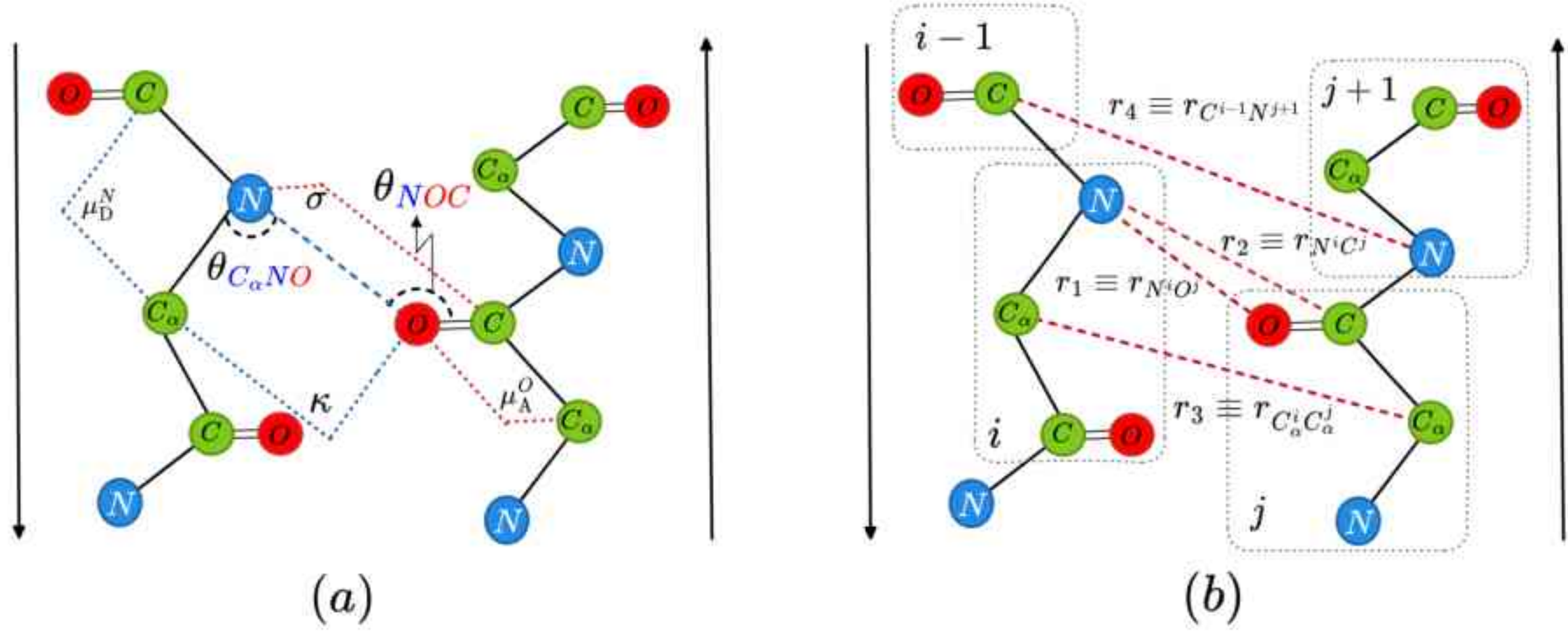
Şekil 4.12 : Alanin amino asitinin omurga azot ${}_A N$ atomu ile omurga oksijen ${}_A O$ atomlarının oluşturdukları $({}_A O_1, {}_A N_2)$ ve $({}_A N_1, {}_A O_2)$ atom çiftleri için farklı d zincir mesafelerinde elde edilen etkileşim enerjilerinin öklid mesafesi r_{12} 'ye göre değişimleri.

davranış farklılığı gösterdiği kolaylıkla anlaşılabacaktır. Şöyle ki; aynı atomlardan oluşan her iki atom çiftinin hidrojen bağı yapma kapasitesinin aynı olması beklenirken; hidrojen bağı oluşturma kriterlerinden birisi olan $r_{ON} < 3.55\text{\AA}$ ve $r_{NO} < 3.55\text{\AA}$ değerlerinde $({}_A O_1, {}_A N_2)$ atom çifti, $({}_A N_1, {}_A O_2)$ atom çiftine göre çok daha fazla gözlemlenmekte ve dolayısıyla çok daha fazla hidrojen bağı yapabilme kapasitesine ve çok daha düşük enerjiye sahiptir (bkz. Şekil.4.12). Bu davranış zincir üzerinde yerel olarak tanımlanan ve $d = 2, 3, 4, 5, 6$ zincir mesafe değerlerinde, diğer tüm amino asit çiftlerine ait donör ve akseptör atom çiftleri için de geçerlidir.

Protein kararlığında önemli rol üstlenen hidrojen bağlarının, hem yönelimsel hem de düzlemsel olma özelliği mevcuttur [84]. Bu iki özelliği kapsayacak transfer edilebilir bir istatistiksel hidrojen bağı enerji fonksiyonu, çoklu parçacık geometrilerini içermelidir. Bu amaçla sadece omurga donör N ve akseptör O atomları arasında oluşan hidrojen bağları ile ilgileneceğiz. Şekil.4.13'te sırasıyla magenta ve mavi renkli kesikli çizgilerle gösterilen $C_\alpha - H \cdots O$ ve siyah renkli $N - H \cdots O$ hidrojen bağları gösterilmiştir.



Şekil 4.13 : β -yapılı proteinlerde omurga atomları arasında oluşan asidik $C_{\alpha}-H \cdots O$ ve $N-H \cdots O$ hidrojen bağları gösterimi.



Şekil 4.14 : Hafif hidrojen atomu tanımlaması içermeyen (a) $N-H \cdots O$ hidrojen bağı oluşum kriterleri ve (b) eğer kriterler yerine getirilmiş ve hidrojen bağı oluşmuşsa ilgili geometrinin kayıt edilmesi ve binlere ayrılabilmesi için tanımlanan dört mesafenin gösterimi.

Geliştireceğimiz ve protein dinamiğine dahil edeceğimiz $N-H\cdots O$ hidrojen bağları için bir mesafe ve dört açı kriteri kullanılmış ve bunların geometrik tanımlamaları Şekil.4.14'de gösterilmiştir. Donor N ve akseptör O atomları sp^2 hibritleşmesine sahiplerdir. Bu nedenle, donor atomunun akseptör düzlemine göre ve akseptör atomunun donor düzlemine göre en ideal oryantasyonları ve ideal durumdan sapmalarının tanımlanması hidrojen bağlarının düzlemsellik özelliğinin yerine getirilmesi bakımından önemlidir. Şekil 4.14.(a)'da $\mu_D^N(C-N-C_\alpha)$ plakası donor düzlemini ve $\mu_A^O(O-C-C_\alpha)$ plakası ise akseptör düzlemini göstermektedir. Akseptör atomunun μ_D^N donor düzlemine göre oryantasyonu; μ_D^N ve $\kappa(C_\alpha-N-O)$ plakalarının normalleri arasındaki açıya göre belirlenmektedir. Diğer bir deyişle, μ_D^N ve κ plakalarının normalleri arasındaki açı için en ideal değer 0° olup, bu değerden $\pm 60^\circ$ 'ye kadarlık sapmalara müsaade edilmektedir. Benzer şekilde, N donor atomunun μ_A^O akseptör düzlemi göre oryantasyonu; μ_A^O plakası ve $\sigma(N-O-C)$ plakalarının normalleri arasındaki açıya göre tayin edilmektedir. Yine bu plakaların normalleri arasındaki açı için de en ideal değer 0° 'dir. Fakat O atomu daha büyük olduğu için ideal değerden $\pm 90^\circ$ 'ye kadar olan sapmalar kabul edilmektedir [84]. Şekil 4.14.(a)'da gösterilen, donor N ve akseptör O atom pozisyonlarında tanımlanan bağ açıları için sırasıyla $\theta_{C_\alpha NO} > 90^\circ$ ve $\theta_{NOC} > 90^\circ$ kriterleri ve donor N ve akseptör O atomları arasındaki mesafe için ise $r_{NO} < 3.6\text{\AA}$ kriteri kullanılmıştır [84]. Bu kriterleri yerine getirerek hidrojen bağı oluşturan ilgili geometriler Şekil 4.14.(b)'de gösterilen dört mesafe kullanılarak kayıt altına alınmakta ve $0.2\text{\AA}$ 'lük binlere ayrılarak istatistiksel potansiyel dizaynında kullanılmaktadır. $N-H\cdots O$ hidrojen bağı geometrisini tanımlamak için kullanılan dört mesafe ise şunlardır: $r_1 \equiv r_{NO}$, $r_2 \equiv r_{NiCj}$, $r_3 \equiv r_{C_\alpha^i C_\alpha^j}$ ve $r_4 \equiv r_{C^{i-1} N^{j+1}}$.

İstatistiksel hidrojen bağı potansiyel enerjisi için tanımladığımız en genel matematiksel ifade Eşitlik.4.10'da verilmiştir.

$$E_{\text{Hb}}^{N\cdots O}(aN, aO, r_1, r_2, r_3, r_4) = -kT \ln \left[1 + \frac{N_{\text{Hb}}^{N\cdots O}(aN, aO, r_1, r_2, r_3, r_4)}{\rho_{\text{Hb}}^{N\cdots O}(aN, aO)} \right] \quad (4.10)$$

Eşitlikteki $N_{\text{Hb}}^{N\cdots O}(aN, aO, r_1, r_2, r_3, r_4)$ terimi hidrojen bağı kriterlerini yerine getirmiş amino asit spesifik (aN, aO) donor akseptör çiftinin r_1, r_2, r_3 ve r_4 mesafe binlerinde toplam gözlemlenme sayısını ifade etmektedir. Eşitlik 4.10'daki 1 değeri; verilen donor akseptör çifti için hidrojen bağı gözlemlenmeyen bin kombinasyonları için matematiksel

olarak tanımsız enerji durumlarını (yani, $-\ln[0]$) engellemek için eklenmiştir.

$$\rho_{\text{Hb}}^{N \cdots O}(aN, aO) = \frac{N_{\text{Hb}}^{N \cdots O}(aN, aO)}{\sum_{i=1}^{20} \sum_{j=1}^{20} N(a_i N, a_j O)} \quad (4.11)$$

Ayrıca, $\rho_{\text{Hb}}^{N \cdots O}(aN, aO)$ terimi ise mol kesri terimi olup (aN, aO) donör akseptör çiftinin yaptığı toplam hidrojen bağı sayısı $N_{\text{Hb}}^{N \cdots O}(aN, aO)$ 'nın, hidrojen bağı yapıp yapmamasından bağımsız protein setindeki tüm donör akseptör çiftlerinin toplam sayısı $\sum_{i=1}^{20} \sum_{j=1}^{20} N(a_i N, a_j O)$ 'na oranı olarak verilmekte ve Eşitlik 4.11 kullanılarak hesaplanmaktadır.

Kullandığımız hidrojen bağı tanımı altında, farklı d zincir mesafelerinde gözlemlenen toplam $N - H \cdots O$ hidrojen bağlarının sayısı Tablo 4.1'de verilmiştir. İkinci kolon (akseptör, donör) çiftleri (yani, akseptör oksijen atomunun N -bitimine daha yakın olduğu durum) ve üçüncü kolon ise (donör, akseptör) çiftlerinin (yani, donör azot atomunun N -bitimine daha yakın olduğu durum) toplam sayılarını göstermektedir. Görüleceği üzere, hidrojen bağı yapan çiftlerin zincir üzerindeki mesafeleri $d \leq 6$ olduğu durumda üçüncü kolondaki (donör, akseptör) çiftlerinin hidrojen bağı yapma eğilimleri, ikinci kolondaki (akseptör, donör) çiftlerinin hidrojen bağı yapma eğilimlerinin yanında çok düşük kalmaktadır. Fakat bu farklılık zincir üzerindeki mesafe $d \geq 7$ olduğu durumda ortadan kalkmaktadır.

	$O \cdots H - N$	$N - H \cdots O$
$i, i+2$	34.444	57
$i, i+3$	1.859.568	4.435
$i, i+4$	1.623.134	4.890
$i, i+5$	37.933	3.429
$i, i+6$	6.597	1.942
$i, i \geq 7$	237.444	207.222

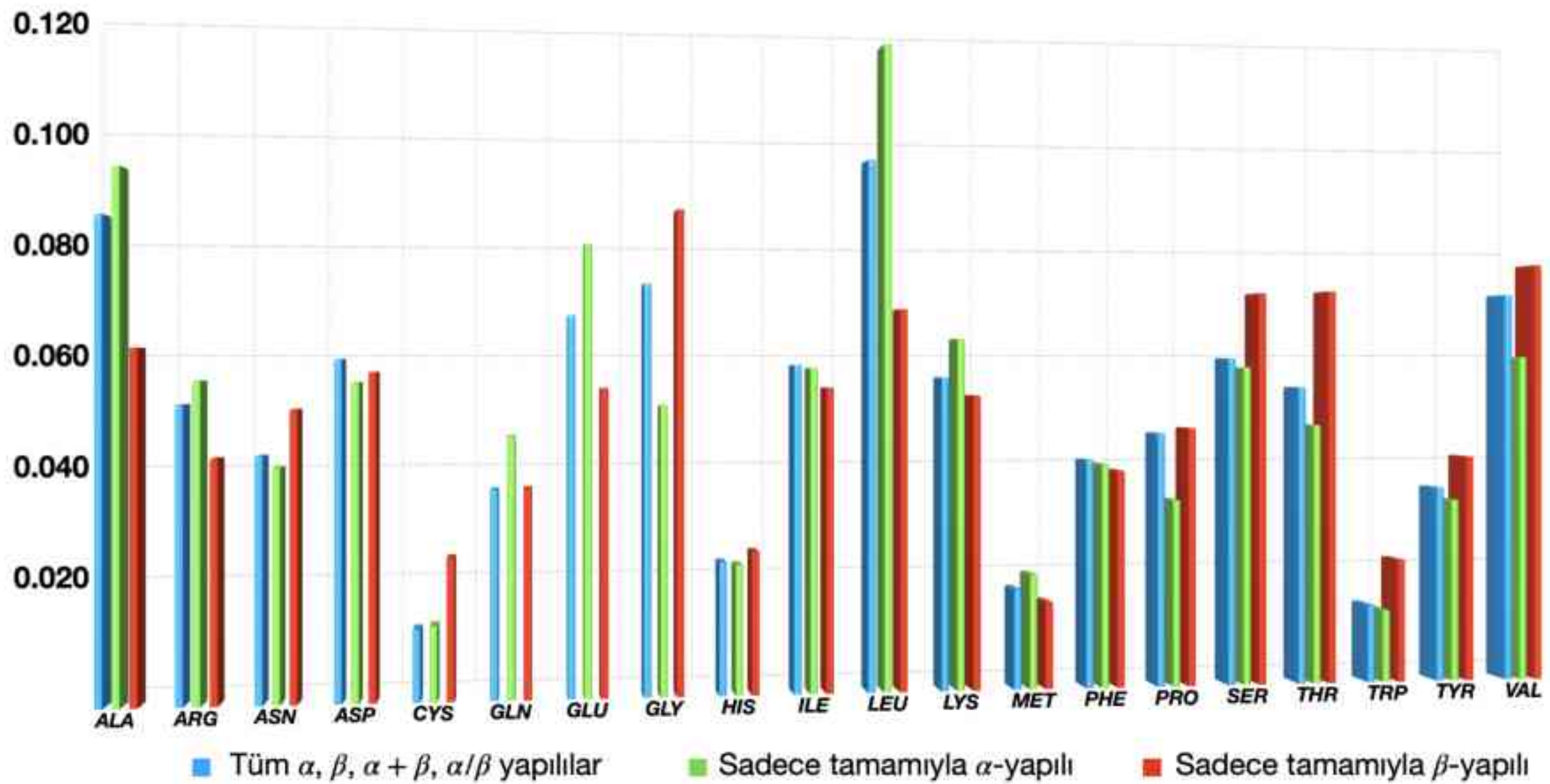
Tablo 4.1 : Dört donör akseptör çifti setleri için farklı d zincir mesafelerinde gözlemlenen toplam hidrojen bağı sayıları.

Hidrojen bağı yapma eğilimindeki bu farklılığı geliştirdiğimiz hidrojen bağı enerji terimine dahil etmek için; hidrojen bağlarını sadece amino asit spesifik değil, aynı zamanda hidrojen bağı yapan donör veya akseptör atomlarından hangisinin protein zincirinin N -bitimine daha yakın olduğu da göz önüne alacak şekilde dizayn edilmiştir.

Dolayısıyla, hidrojen bağı enerji terimimiz, hidrojen bağı yapabilecek $2 \times 20 \times 20 = 800$ farklı donör ve akseptör çifti kombinasyonunu içermektedir.

4.2.3 Protein zinciri üzerinde yerel olan amino asit üçlüleri ve dörtlüleri için transfer edilebilir istatistiksel etkileşim enerji terimleri dizaynı

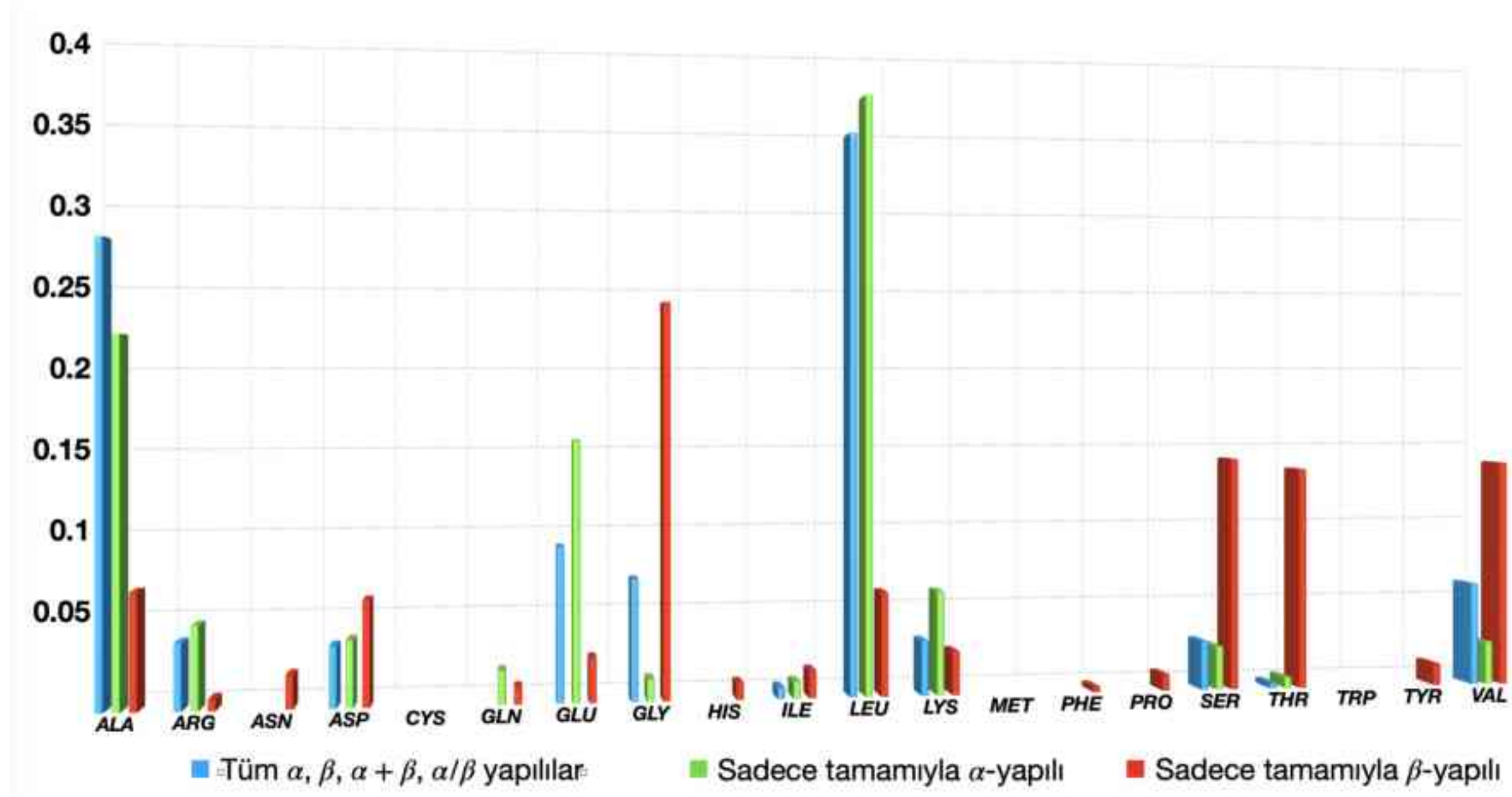
Bu tez kapsamında yaptığımız moleküler dinamik çalışmasında; zincir üzerinde yerel olan açısall etkileşme terimlerinin, protein termodinamiği ve kinetiğini deneysel gözlemlerle uyumlu olacak şekilde betimlenmesinde ne kadar kritik ve önemli rol oynadığını göstermiştik. Benzer katkının tüm atomların temsil edildiği, komplekslik ve karmaşıklık seviyesi çok daha yüksek olan modellerde de sağlayacağı düşüncesiyle; protein zinciri üzerinde ardışık olan amino asit üçlüleri (a_1, a_2, a_3) ve dörtlüleri (a_1, a_2, a_3, a_4) (bkz. Şekil 4.10) ele alınmıştır. Bu bağlamda, amino asit üçlüleri ve dörtlülerine mensup amino asitlerin tüm omurga (N , C_α , C ve O) atomları ve sadece kenar zincir C_β atomları arasında mesafeler, bağ açıları ve dihedral açıları tanımlayarak; amino asit üçlü ve dörtlülerinde ki atomların geometrik tercihlerini çok atomlu ve çok parametrelili enerji terimlerine dönüştürülmesi ve böylelikle protein dinamiğine dahil edilmesi planlanmıştır.



Şekil 4.15 : Tüm farklı protein katlı yapılarını, sadece tamamıyla α -yapılarını ve sadece tamamıyla β -yapılarını içeren üç farklı protein setinde amino asitlerin gözlemlenme sıklığının gösterimi.

Amaçlanan şey ise; (i) protein katlı yapılarının en düşük enerjili yapı olmasını, (ii) protein konformasyon uzayının efektif bir şekilde örneklenmesini, (iii) tuzak oluşturabilecek

konformasyonların ziyaret edilmemesini, ve (iv) yerel konformasyonel tercihlerin doğru bir şekilde betimlenmesini ve bunların protein dinamiğine katkılarının ince ayar yapılabilecek şekilde tasarlanabilmesini sağlamaktır.



Şekil 4.16 : Tüm farklı protein katlı yapılarını, sadece tamamiyla α -yapılarını ve sadece tamamiyla β -yapılarını içeren üç farklı protein setinde en sık gözlemlenen ilk yüz amino asit üçlüsünde, amino asitlerin gözlemlenme sıklığının gösterimi.

Literatürde protein katlı yapı analizleri ile amino asitlerin gösterdikleri farklı eğilimler (mesela, hangi amino asitin hangi ikincil yapılarda bulunmayı tercih ettiği vd.) detaylı olarak incelenmiştir [85–88]. Benzer analizi kullandığımız 25,801 proteinden oluşan protein setimizde ki; (i) tüm proteinleri, aynı protein setinde bulunan (ii) sadece α -yapılı proteinler, ve (iii) sadece β -yapılı proteinlerden oluşan protein setleri için bu tez çalışmasında da yapılmıştır.

Bir amino asitin farklı protein aile setlerindeki gözlemlenme sıklığı P_a^i , Eşitlik 4.12’de verilen matematiksel form ile tanımlanmaktadır.

$$P_a^i = \frac{N_a^i}{\sum_a N_a^i} \quad (4.12)$$

Eşitlik 4.12’deki a alt indisi amino asit türünü, i üst indisi protein ailesini, N_a^i terimi i protein ailesinde gözlemlenen toplam a amino asit türü sayısını ve $\sum_a N_a^i$ terimi i ailesi protein setindeki toplam amino asit sayısını göstermektedir. Farklı protein ailelerinden elde edilen P_a^i değerlerinde gözlemlenen farklılığın kıyaslaması, Eşitlik 4.13’te verilen Γ_a^{ij}

parametresi kullanılarak yapılmıştır.

$$\Gamma_a^{ij} = \frac{|P_a^i - P_a^j|}{\min(P_a^i, P_a^j)} \quad (4.13)$$

Eşitlik 4.13'teki i ve j üst indisleri protein ailelerini temsil etmektedir.

Şekil 4.15 dikkatli incelendiğinde; LEU, ALA, GLU amino asitleri tamamıyla α -yapılı proteinlerde ve GLY, VAL, THR amino asitleri tamamıyla β -yapılı proteinlerde daha yüksek gözlemlenme sıklığına sahiptirler. Aynı amino asitlerin iki protein ailesindeki gözlemlenme sıklıklarındaki farklılıkları Γ_a^{ij} parametresini kullanarak incelediğimizde; $\Gamma_{LEU}^{\alpha,\beta} = 0.69$, $\Gamma_{ALA}^{\alpha,\beta} = 0.50$, $\Gamma_{GLU}^{\alpha,\beta} = 0.47$, $\Gamma_{GLY}^{\alpha,\beta} = 0.68$, $\Gamma_{VAL}^{\alpha,\beta} = 0.29$ ve $\Gamma_{THR}^{\alpha,\beta} = 0.52$ değerleri elde edilmektedir.

Diğer taraftan, protein zinciri üzerinde ardışık amino asit üçlülerinin aynı protein setlerindeki gözlemlenme sıklığı da incelenmiştir. Bu analiz sırasında 8,000 farklı amino asit üçlüsünün tümünü ele almak yerine; tamamıyla α -yapılı ve tamamıyla β -yapılı proteinlerde en sık gözlemlenen ilk 100 amino asit üçlülerini üzerinden, amino asit gözlemlenme sıklıkları hesaplanmıştır. Ayrıca en sık gözlemlenen ilk 10 amino asit üçlülerini Tablo 4.2'de verilmiştir. Yapılan analizin Şekil 4.16'da gösterilen sonuçları dikkate değerdir. En sık gözlemlenen ilk 100 amino asit üçlülerini üzerinden yapılan gözlemlenme sıklığı farkları analizinde elde edilen değerler; $\Gamma_{LEU}^{\alpha,\beta} = 4.79$, $\Gamma_{ALA}^{\alpha,\beta} = 2.14$, $\Gamma_{GLU}^{\alpha,\beta} = 4.87$, $\Gamma_{GLY}^{\alpha,\beta} = 17.00$, $\Gamma_{VAL}^{\alpha,\beta} = 4.25$ ve $\Gamma_{THR}^{\alpha,\beta} = 19.50$ 'dir.

Tablo 4.2 : Tüm farklı ailelere mensup proteinlerden, tamamıyla α -yapılı ve tamamıyla β -yapılı proteinlerden oluşturulan üç farklı protein setinde en sık gözlemlenen ilk on amino asit üçlülerinin gösterimi.

	Tüm α , β , $\alpha + \beta$, α/β yapılı	Tamamıyla α -yapılı	Tamamıyla β -yapılı
1	ALA-ALA-ALA	ALA-ALA-ALA	SER-GLY-SER
2	ALA-LEU-ALA	ALA-ALA-LEU	ASP-GLY-THR
3	ALA-ALA-LEU	GLU-LEU-LEU	PRO-ASP-GLY
4	LEU-ALA-ALA	LEU-LEU-GLU	VAL-SER-GLY
5	GLU-ALA-LEU	ALA-LEU-ALA	SER-VAL-THR
6	LEU-LEU-ALA	LEU-LEU-ALA	VAL-LYS-VAL
7	ALA-LEU-LEU	LEU-ALA-ALA	THR-LEU-THR
8	GLU-LEU-LEU	GLU-GLU-LEU	VAL-VAL-SER
9	LEU-LEU-GLU	ALA-LEU-LEU	VAL-THR-GLY
10	LEU-ALA-GLU	GLU-ALA-LEU	SER-GLY-GLY

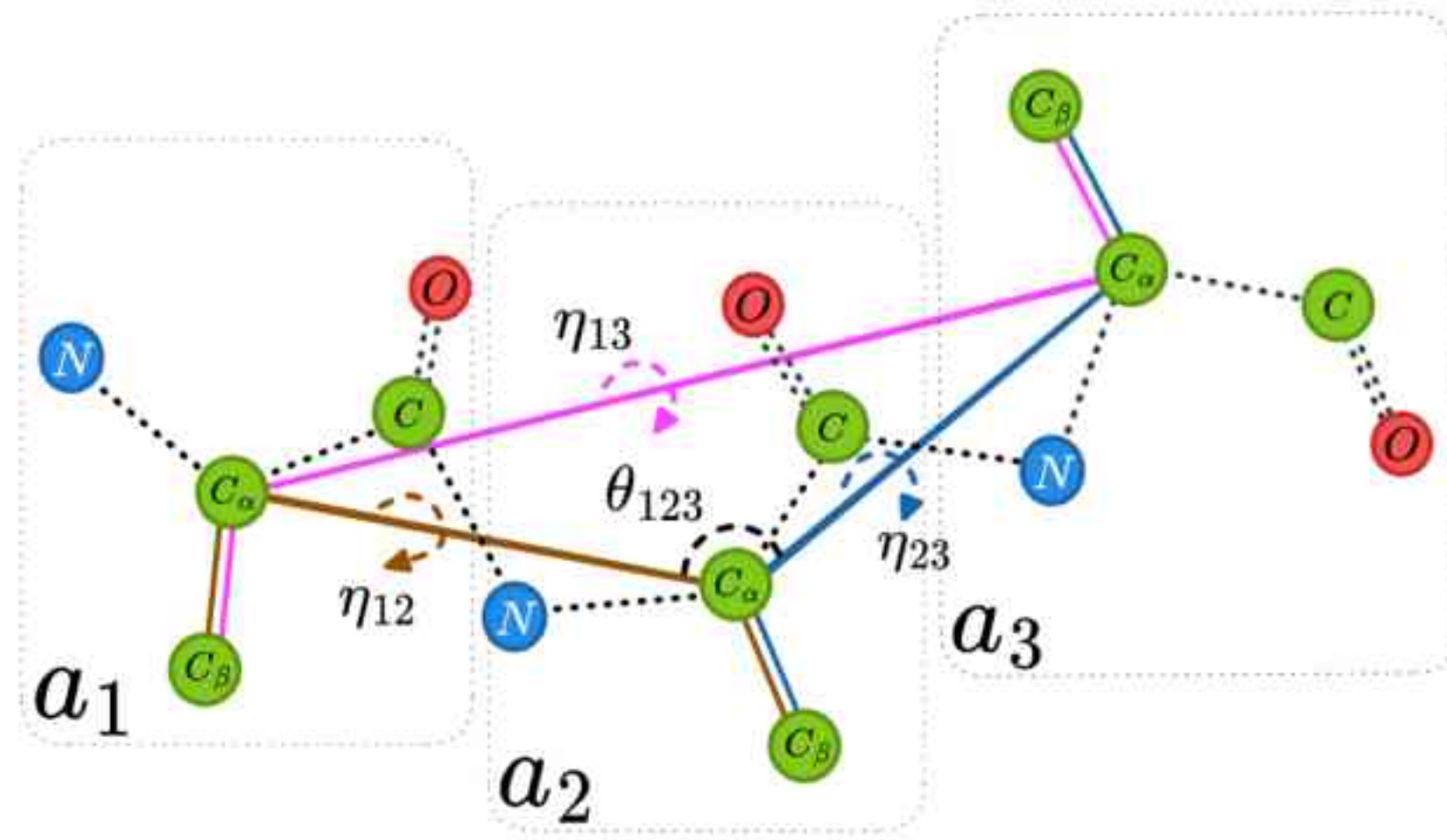
Bu gözlemlerden elde edilen sonuçlar şunları ifade etmektedir: tüm amino asitler, tüm farklı protein ailesine mensup proteinlerde kıyaslanabilir farklılıklar ile gözlemlenseler dahi; bu ailelere mensup proteinlerde geliş güzel bir şekilde değil, tam aksine her protein ailesinde belirli amino asit üçlülerinin içinde çok daha yüksek sıklıkla olacak şekilde bulunmayı tercih etmektedirler. Diğer bir deyişle, protein zinciri üzerinde ardışık amino asitlerin oluşturduğu üçlülerin; proteinlerin katlı yapılarının seçilmesinde veya oluşmasında belirleyici rolü olduğunu işaret etmektedir. Bu nedenle, geliştirdiğimiz amino asit üçlüsü enerji terimlerinin protein dinamiğinde de belirleyici rol üstlenmesi beklenmektedir.

4.2.3.1 $C_\alpha - C_\beta$ Kovalent Bağlarının Birbirine Göre Hizalanmasını η Dihedral ve θ Bağ Açıları ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\eta-\theta}$

Protein zinciri üzerinde ardışık amino asitlerin oluşturduğu her amino asit üçlüsündeki (a_1, a_2, a_3) atomların geometrik tercihlerini belirlemek için tanımladığımız; bağ ve dihedral açıları ya üç amino asitin tamamında ya da (a_1, a_2) , (a_2, a_3) ve (a_1, a_3) şeklinde ifade edilen üç tane amino asit çiftinin atomları arasında tanımlanmaktadır. Amino asit üçlüsünden türetilen amino asit çiftleri ve açılarının numaralandırılması veya indekslenmesi, (a_1, a_2, a_3) gösteriminde olduğu gibi üçlü içindeki sıralarına göre yapılmaktadır.

Amino asit üçlülerindeki amino asit çiftlerindeki $C_\alpha - C_\beta$ kovalent bağlarının birbirlerine göre hizalanmalarını $\eta_{12} \equiv \angle C_\beta^1 C_\alpha^1 C_\alpha^2 C_\beta^2$, $\eta_{23} \equiv \angle C_\beta^2 C_\alpha^2 C_\alpha^3 C_\beta^3$ ve $\eta_{13} \equiv \angle C_\beta^1 C_\alpha^1 C_\alpha^3 C_\beta^3$ dihedral açıları ve C_α atomları arasında tanımlanan $\theta_{123} \equiv \angle C_\alpha^1 C_\alpha^2 C_\alpha^3$ bağ açısı ile tasvir edilmektedir. Bu açılar Şekil.4.17’de grafiksel olarak gösterilmiştir. η dihedral açılarının alt indisleri, bu açıların verilen amino asit üçlüsündeki hangi amino asit ikilisine ait olduğunu belirtmektedir. Görüleceği üzere, mesela η_{12} dihedral açısı C_β^1 , C_α^1 , C_α^2 ve C_β^2 atomları arasında tanımlanmakta ve $\eta_{12} \equiv \angle C_\beta^1 C_\alpha^1 C_\alpha^2 C_\beta^2$ şeklinde ifade edilmektedir. Benzer şekilde, θ_{123} bağ açısı C_α^1 , C_α^2 ve C_α^3 atomları arasında tanımlanmakta ve matematiksel olarak $\theta_{123} \equiv \angle C_\alpha^1 C_\alpha^2 C_\alpha^3$ şeklinde gösterilmektedir. Bu tanımlama ve gösterim şekli tüm açısal terimler için aynıdır.

$[-180^\circ, 180^\circ]$ arasında değerler alan dihedral açıları 20° ’lik ve $[0^\circ, 180^\circ]$ arasında değerler alan bağ açıları ise 10° ’lik eşit parçalara bölünerek; her biri için 18 tane ayrık aralık/bin oluşturulmuştur. Protein modelimizde kullanılmak üzere amino asit üçlülerinin içerdikleri atomların birbirlerine göre konum veya geometrik dizilim tercihlerine dayalı olarak



Şekil 4.17 : Protein zinciri üzerinde ardışık olan (a_1, a_2, a_3) amino asit üçlüsündeki, üç amino asit çiftlerinde $C_\alpha - C_\beta$ kovalent bağlarının birbirlerine göre hizalanmalarını tasvir etmek için tanımlanan η_{12} , η_{23} , η_{13} dihedral açıları ve θ_{123} bağ açısı terimlerinin gösterimi.

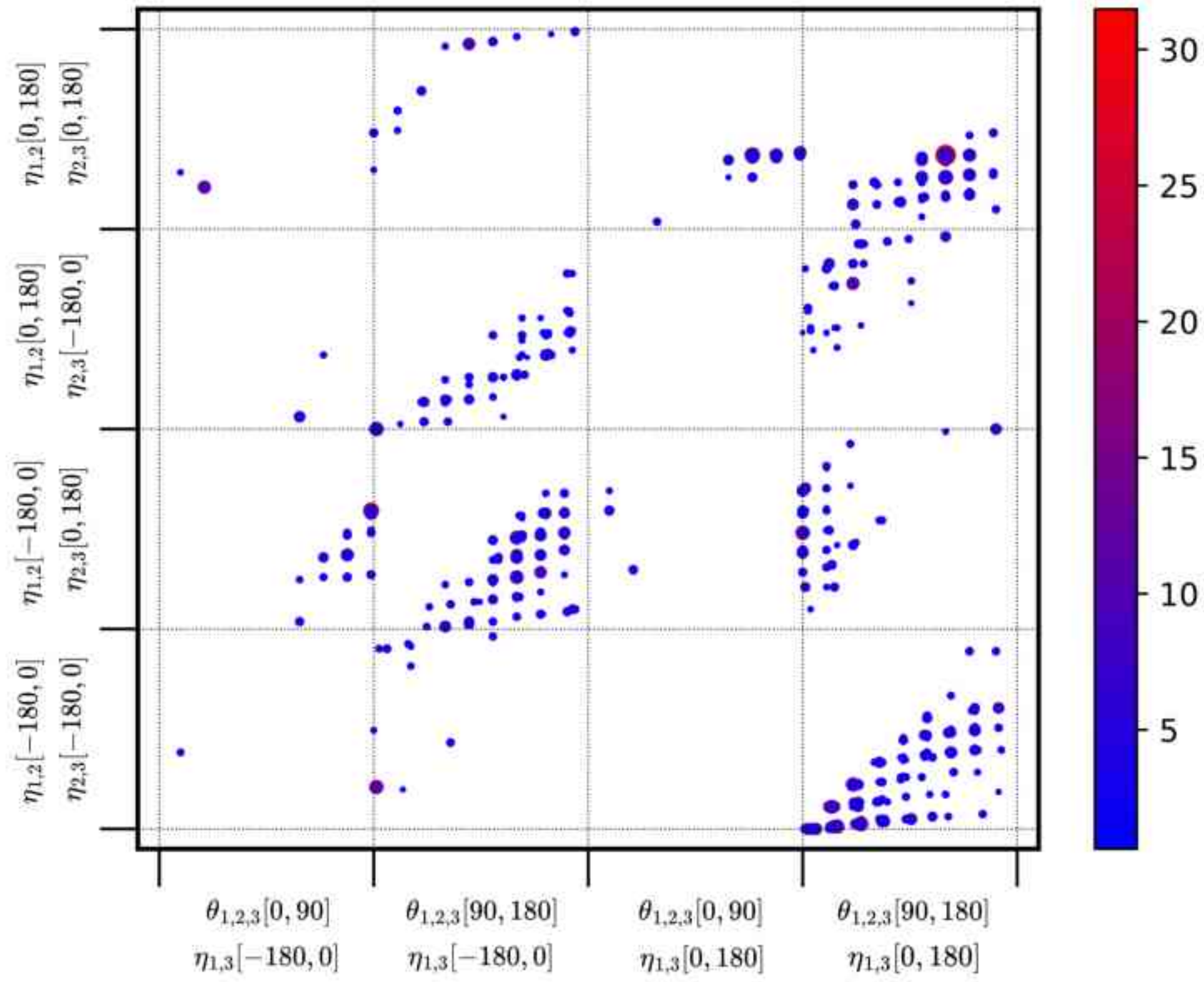
geliştirilen transfer edilebilir istatistiksel enerji fonksiyonları; Eşitlik.4.14’de gösterildiği gibi $20 \times 20 \times 20 = 8,000$ tane farklı amino asit üçlüsü ve dört tane açı parametresi kullanılarak oluşturulmuştur. Her bir amino asit üçlüsü için toplam $18 \times 18 \times 18 \times 18 = 104,976$ tane içine girilebilir durum söz konusudur. Dolayısıyla, her bir enerji fonksiyonu $8,000 \times 104,976 \simeq 84 \times 10^6$ tane içine girilebilir farklı mikroskopik durumu içermektedir. Belirtmek gerekir ki; geliştirilen her bir amino asit üçlüsü enerji fonksiyonu bu kadar yüksek sayıda mikroskopik durumu içermesine rağmen, her proteinde tüm amino asit üçlüsü çeşitlerinin bulunması olası değildir. Yani, proteinler içerdikleri amino asit üçlüsü çeşidi sayısına bağlı olarak, içine girilebilecek durum sayısı farklılık gösterecektir.

$$E_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3), \eta_{12}, \eta_{23}, \eta_{13}, \theta_{123}) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3), \eta_{12}, \eta_{23}, \eta_{13}, \theta_{123})}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.14)$$

Eşitlik.4.14’de $N_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3), \eta_{12}, \eta_{23}, \eta_{13}, \theta_{123})$ terimi, η_{12} , η_{23} , η_{13} dihedral ve θ_{123} bağ açısı bin değerleri kombinasyonunda gözlemlenen toplam (a_1, a_2, a_3) amino asit üçlüsünün sayısını ve $\rho_{\text{üçlü}}(a_1, a_2, a_3)$ terimi ise (a_1, a_2, a_3) amino asit üçlüsünün mol kesrini belirtmekte ve değeri Eşitlik.4.15 kullanılarak hesaplanmaktadır.

$$\rho_{\text{üçlü}}(a_1, a_2, a_3) = \frac{N_{\text{üçlü}}(a_1, a_2, a_3)}{\sum_{i=1}^{20} \sum_{j=1}^{20} \sum_{k=1}^{20} N_{\text{üçlü}}(a_1^i, a_2^j, a_3^k)} \quad (4.15)$$

Eşitlik.4.15’de $N_{\text{üçlü}}(a_1, a_2, a_3)$ terimi (a_1, a_2, a_3) amino asit üçlüsünün protein setindeki toplam gözlemlenme sayısını ve payda terimi ise mümkün olan tüm 8.000 amino asit üçlüsünün tamamının toplam gözlemlenme sayısını ifade etmektedir. Payda daki toplam terimlerindeki üst limit olarak verilen 20 sayısı, toplam amino asit çeşidi sayısını vermektedir. Eşitlik.4.15’deki bu iki terimin oranı; kuvazi-kimyasal yaklaşımda mol kesri olarak tanımlanmakta ve referans durumunu temsil etmektedir [89]. $\rho_{\text{üçlü}}(a_1, a_2, a_3)$ mol kesri yani referans durum değeri, amino asit üçlüleri için geliştirilen tüm enerji terimleri için aynıdır.



Şekil 4.18 : Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3), \eta_{12}, \eta_{23}, \eta_{13}, \theta_{123})$ enerji fonksiyonunun minimum değere sahip olduğu η_{12} , η_{23} , η_{13} ve θ_{123} parametrelerinin bin kombinasyonlarının gösterimi.

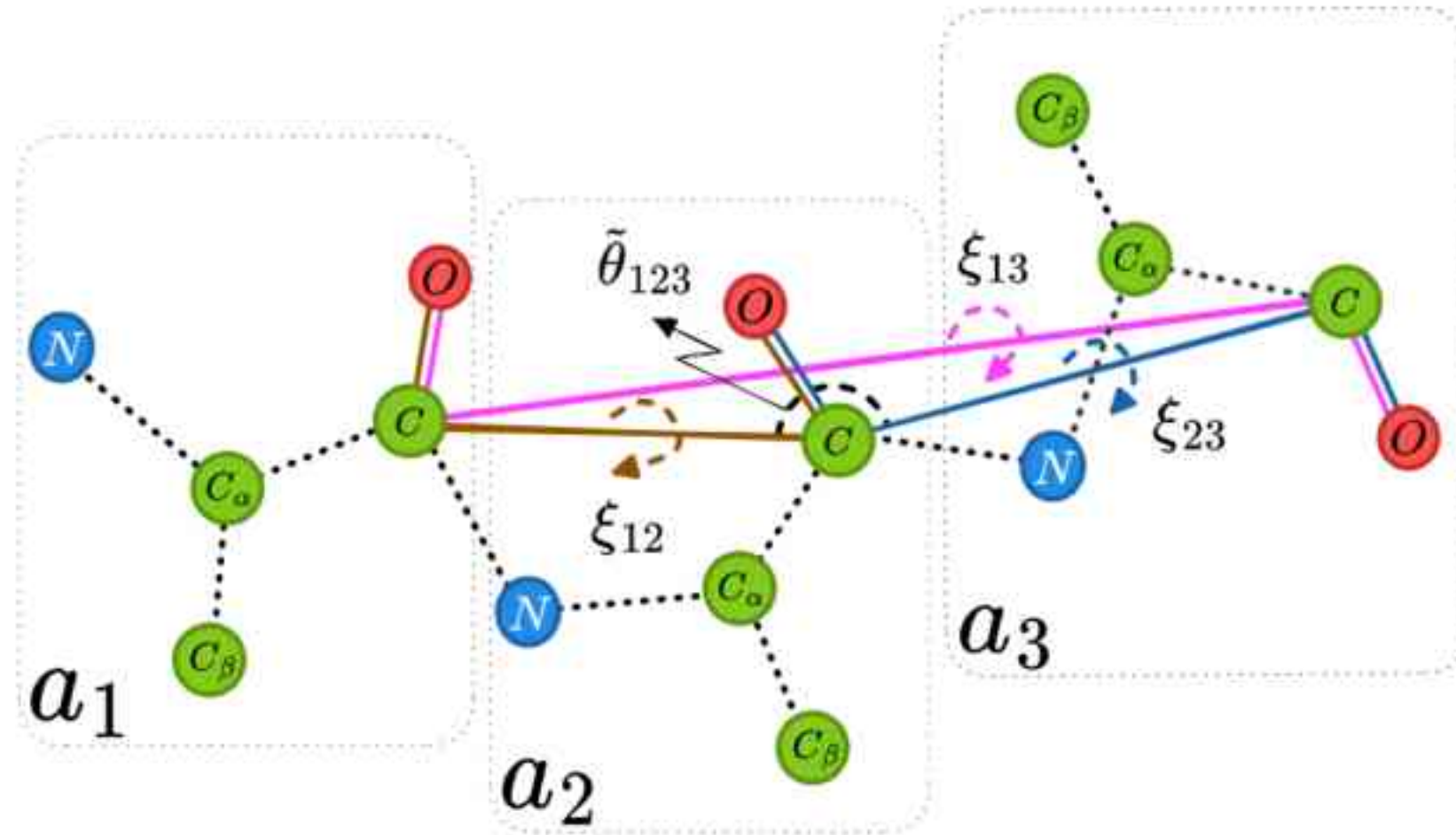
Eşitlik 4.14 ve geliştirdiğimiz diğer amino asit üçlüleri enerji terimlerinde; yukarıda ifade edildiği her amino asit üçlüsü için dört boyutlu geometrik tanımlamalar yapılmaktadır. Şekil 4.18’de 25,801 proteinden oluşan protein setimiz kullanılarak; her amino asit üçlüsünde, bu dört parametrenin hangi bin kombinasyonunun en sık gözlemlendiği veya minimum enerjiyi verdiği değerler gösterilmektedir. Şekil 4.18’de 4-boyutlu enerji fonksiyonumuzun, 2-boyuta projeksiyonu; x-ekseni ve y-ekseninde gösterilen

ikişer bağımsız parametrenin yeniden indekslenerek, tek boyuta indirgenmesi şeklinde gerçekleştirilmiştir.

Şekil 4.18’de gösterilen noktalar gözlemlenme sıklığına veya enerjinin düşüklüğüne bağlı olarak farklı büyüklük ve renk skalası kullanılarak gösterilmişlerdir. Yani, gözlemlenme sıklığı arttıkça hem noktanın büyüklüğü artıyor hem de maviden kırmızıya doğru rengi değişiyor.

Belirtilmesi gerekir ki; 8,000 amino üçlüsünün en düşük enerji değerlerine sahip olduğu parametre bin kombinasyonlarının gösterildiği Şekil 4.18’de sadece 438 data noktası ($438/8000 = 0.055$) bulunmaktadır. Diğer bir deyişle, tüm amino asit üçlülerinin %95’i, 438 farklı parametre bin kombinasyonlarından birisinde minimum enerjiye sahiptir. Yani, amino asit üçlülerinin $E_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3), \eta_{12}, \eta_{23}, \eta_{13}, \theta_{123})$ enerji teriminden elde edilen minimum enerji uzayı çok kısıtlı olup, bu modelleme çalışmalarında istenen ve aranan durumdur.

4.2.3.2 $C = O$ Kovalent Bağlarının Birbirine Göre Hizalanmasını ξ Dihedral ve $\tilde{\theta}$ Bağ Açıları ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\xi-\tilde{\theta}}$

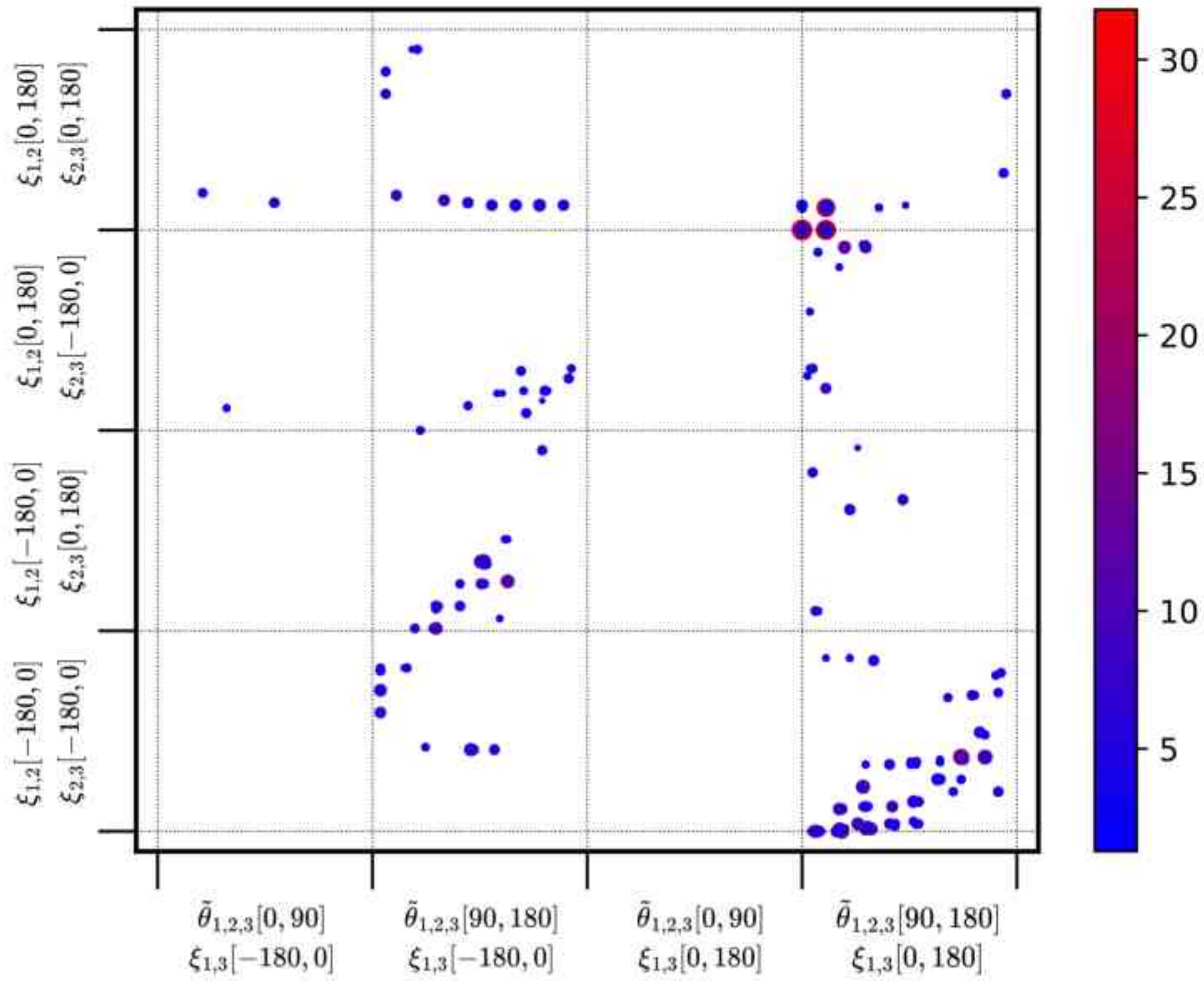


Şekil 4.19 : Her (a_1, a_2, a_3) amino asit üçlüsündeki üç amino asit çiftindeki $C = O$ kovalent bağlarının birbirlerine göre hizalanmaları ve geometrik tercihlerini belirlemek için kullanılan ξ_{12} , ξ_{23} , ξ_{13} dihedral açıları ve $\tilde{\theta}_{123}$ eğilme açı terimlerinin gösterimi.

$E_{\text{üçlü}}^{\xi-\tilde{\theta}}$ istatistiksel enerji terimi, $E_{\text{üçlü}}^{\eta-\theta}$ enerji terimi ile benzer olup, sadece $C_{\alpha} - C_{\beta}$ kovalent bağları yerine $C = O$ kovalent bağlarının birbirlerine göre hizalanmasına odaklanmaktadır. Bu amaçla grafiksel gösterimi Şekil 4.19’da yapılan $\xi_{12} \equiv \angle O^1 C^1 C^2 O^2$, $\xi_{23} \equiv \angle O^2 C^2 C^3 O^3$, $\xi_{13} \equiv \angle O^1 C^1 C^3 O^3$ dihedral açıları ve $\tilde{\theta}_{123} \equiv \angle C^1 C^2 C^3$ bağ açısı terimleri tanımlanmıştır.

$$E_{\text{üçlü}}^{\xi-\tilde{\theta}} \left((a_1, a_2, a_3), \xi_{12}, \xi_{23}, \xi_{13}, \tilde{\theta}_{123} \right) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^{\xi-\tilde{\theta}} \left((a_1, a_2, a_3), \xi_{12}, \xi_{23}, \xi_{13}, \tilde{\theta}_{123} \right)}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.16)$$

Eşitlik 4.16’de matematiksel formu verilen $E_{\text{üçlü}}^{\xi-\tilde{\theta}} \left((a_1, a_2, a_3), \xi_{12}, \xi_{23}, \xi_{13}, \tilde{\theta}_{123} \right)$ enerji fonksiyonunun, tüm amino asit üçlüleri için minimum enerji değerine sahip olduğu parametre bin kombinasyonları Şekil 4.20’te gösterilmiştir.

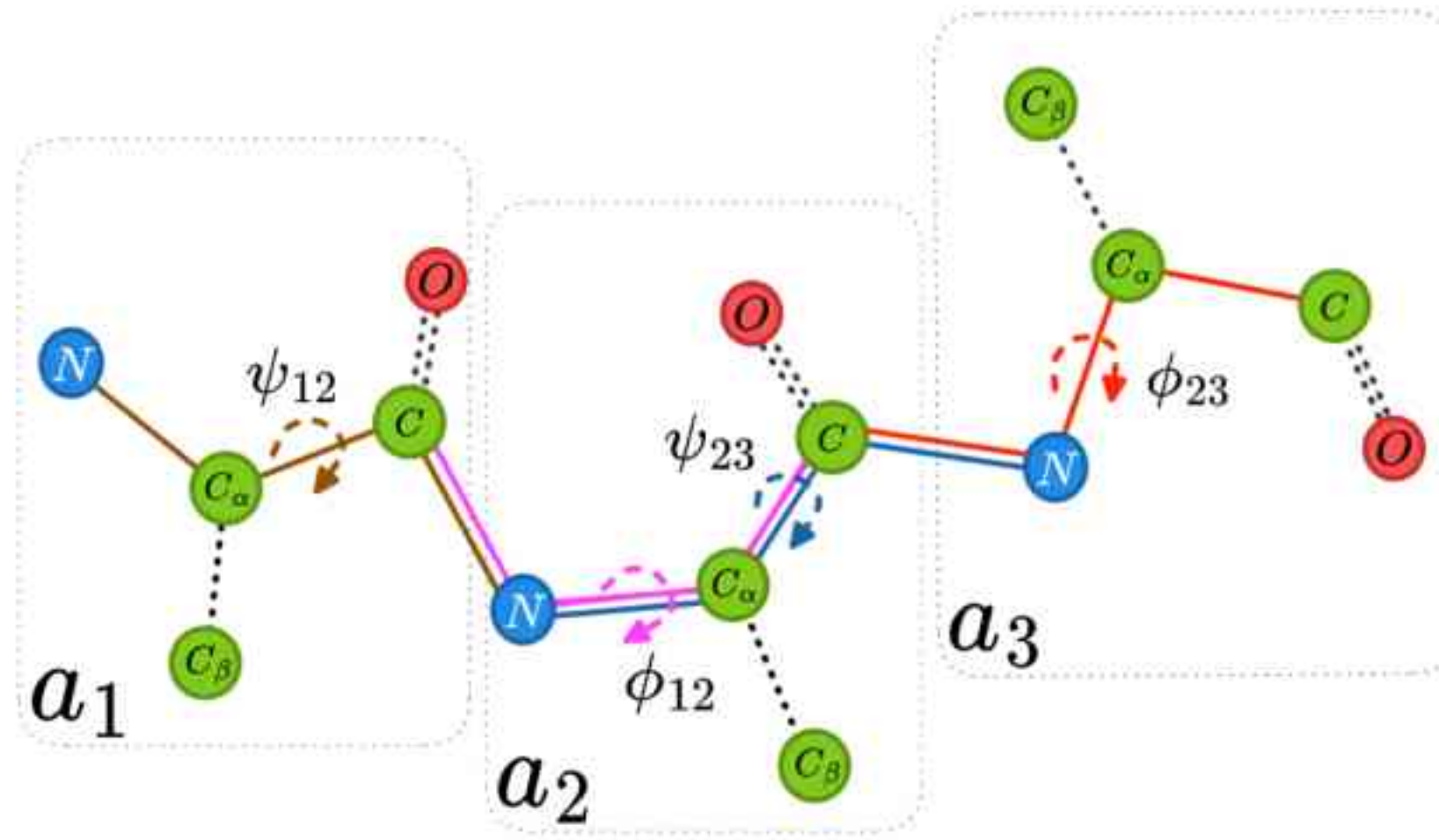


Şekil 4.20 : Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\xi-\tilde{\theta}} \left((a_1, a_2, a_3), \xi_{12}, \xi_{23}, \xi_{13}, \tilde{\theta}_{123} \right)$ enerji fonksiyonunun minimum değere sahip olduğu ξ_{12} , ξ_{23} , ξ_{13} ve $\tilde{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi

8,000 amino asit üçlüsü verisini içeren Şekil 4.20’de sadece 151 farklı parametre bin kombinasyonu gözlemlenmiştir. Yani, tüm amino asit üçlülerinin %98’i, sadece 151 farklı parametre bin kombinasyonlarından birisinde minimum enerjiye sahip olmaktadır.

4.2.3.3 Amino asit üçlüsünde tanımlanan ψ ve ϕ Dihedral Açılar ile Tasvir Edilen Enerji Terimi, $E_{\text{üçlü}}^{\psi-\phi}$

Proteinlerin katlı yapılarında amino asitlerin aldıkları ψ ve ϕ dihedral açı değerleri, protein yapı analizinde, tayininde ve modellemesinde önemli rol oynamaktadır. 1963 yılında Ramachandran tarafından ortaya konan Ramachandran grafiği [90]; günümüzde bir çok deneysel (X-ışını kristalografisi ve NMR) ve protein modelleme yaklaşımında (mesela, homoloji) elde edilen protein yapılarının test edilmesi yaygın bir şekilde kullanılmaktadır.

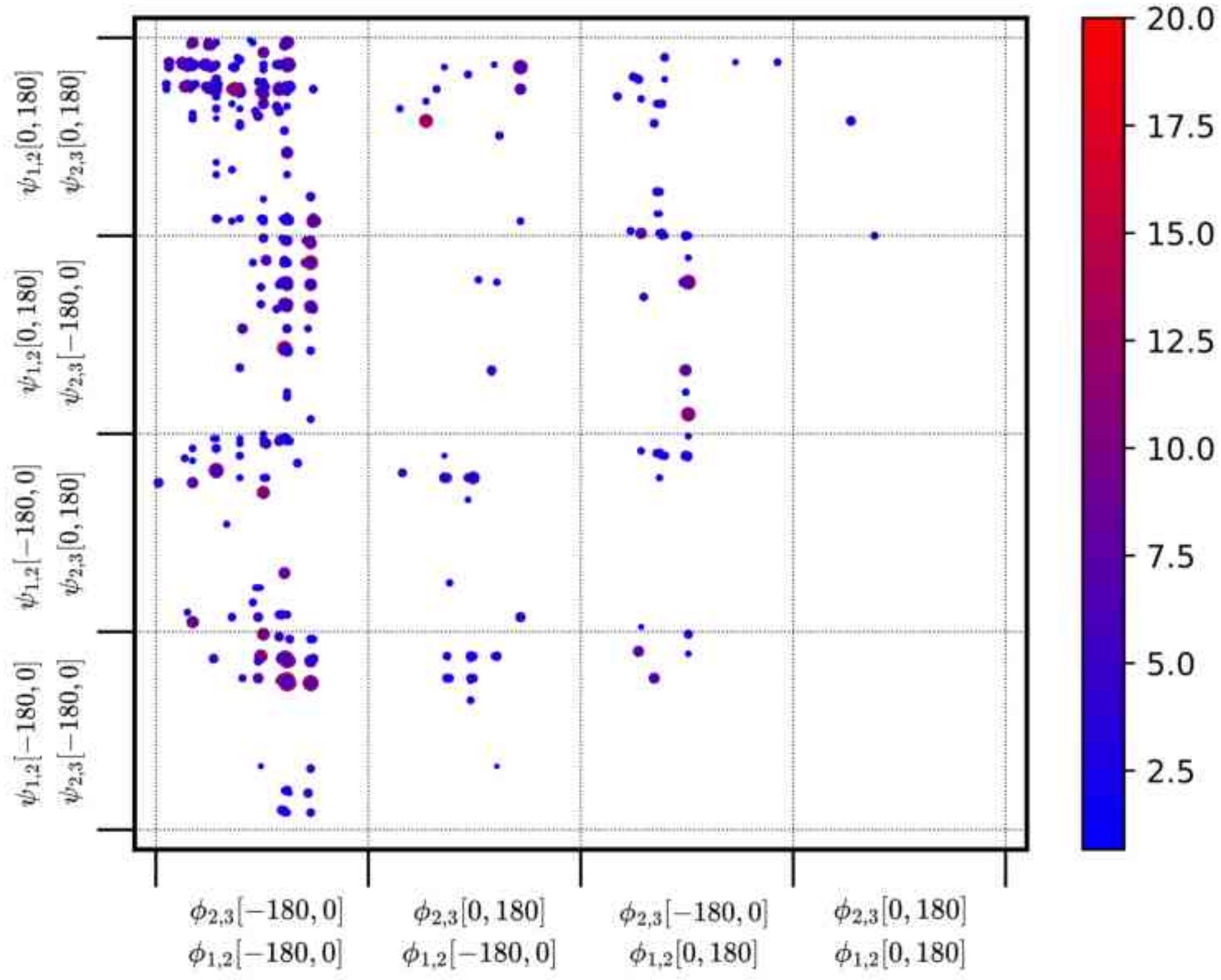


Şekil 4.21 : Her (a_1, a_2, a_3) amino asit üçlüsünde tanımlanan ψ_{12} , ϕ_{12} , ψ_{23} ve ϕ_{23} dihedral açılarının grafiksel gösterimi.

Proteinlerin katlı yapılarındaki konformasyonel tercihler hakkında en güçlü sinyalleri/öbeklenmeleri veren ψ ve ϕ açıları kullanılarak amino asit üçlüleri için yeni bir istatistiksel enerji terimi tanımlanmıştır. Burada ψ ve ϕ açılarını tek başına amino asitler için ele almak yerine amino asit üçlüleri içinde tanımlamak istememizin sebebi; komşu amino asit spesifikliği veya seçiciliğini istatistiksel enerji terimimize dahil etmektir.

$$E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23}) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.17)$$

Bu bağlamda, grafiksel gösterimi Şekil 4.21’de yapılan ve $\psi_{12} \equiv \angle N^1 C_\alpha^1 C^1 N^2$, $\phi_{12} \equiv \angle C^1 N^2 C_\alpha^2 C^2$, $\psi_{23} \equiv \angle N^2 C_\alpha^2 C^2 N^3$ ve $\phi_{23} \equiv \angle C^2 N^3 C_\alpha^3 C^3$ şeklinde tanımlanan dihedral açıların fonksiyonu olan $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ istatistiksel enerjimizin matematiksel formu Eşitlik 4.17’de verilmiştir.

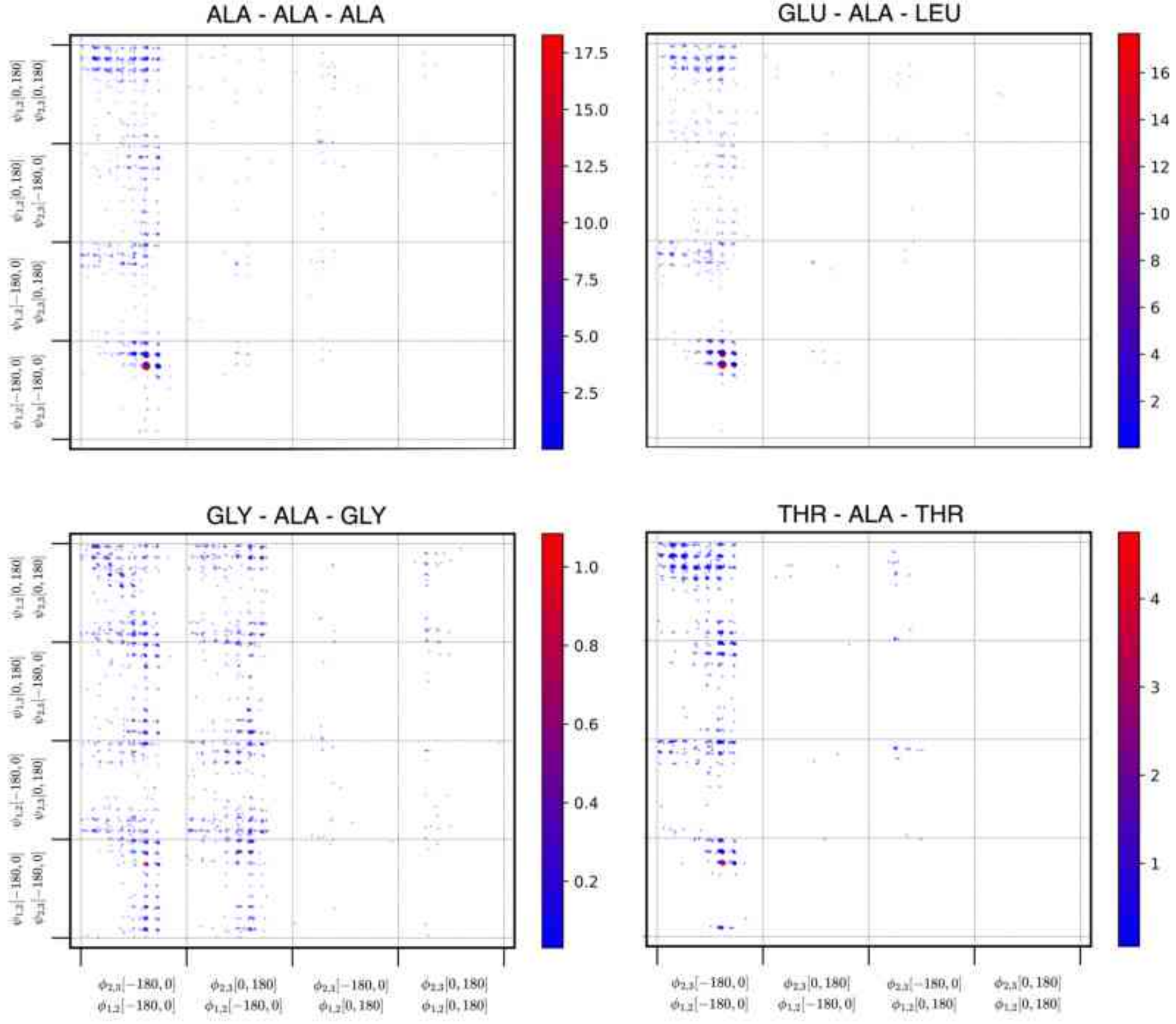


Şekil 4.22 : Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerji fonksiyonunun minimum değere sahip olduğu $\psi_{12}, \phi_{12}, \psi_{23}$ ve ϕ_{23} parametrelerinin bin kombinasyonlarının gösterimi

8.000 amino asit üçlüsü verisini içeren Şekil 4.22’de sadece 360 farklı parametre bin kombinasyonu gözlemlenmiştir. Bu bin kombinasyonu sayısı $E_{\text{üçlü}}^{\eta-\theta}$ ve $E_{\text{üçlü}}^{\xi-\tilde{\theta}}$ enerji fonksiyonlarından elde edilen rakamlardan daha yüksektir. Bunun sebebi ise; glisin amino asitinin hiç kenar zincir atomu olmaması ve prolinin kenar zincirinin halka yaparak omurga N atomuna bağlı halka oluşturması dolayısıyla, bu amino asitlerinin ψ ve ϕ açılarının Ramachandran grafiğinde yasak olan değerleri alabilmeleridir. GLY ve PRO içeren amino asit üçlüleri data setimizden çıkarıldığında gözlemlenen bin kombinasyonu sayısı 360’dan 133’e düşmektedir.

Diğer taraftan amino asit üçlü merkezinde aynı amino asit (ALA) seçildiğinde, tamamıyla α -yapılı proteinlerde en sık gözlemlenen (ALA, ALA, ALA) ve (GLU, ALA, LEU) amino asit üçlüleri ile tamamıyla β -yapılı proteinlerde en sık gözlemlenen (GLY, ALA, GLY) ve (THR, ALA, THR) amino asit üçlüleri ayrı ayrı ele alınarak incelenmiştir. Verilen amino asit üçlüleri için tüm parametre bin kombinasyonlarında

$E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerji fonksiyonunun aldığı değerler Şekil 4.23'te gösterilmiştir.



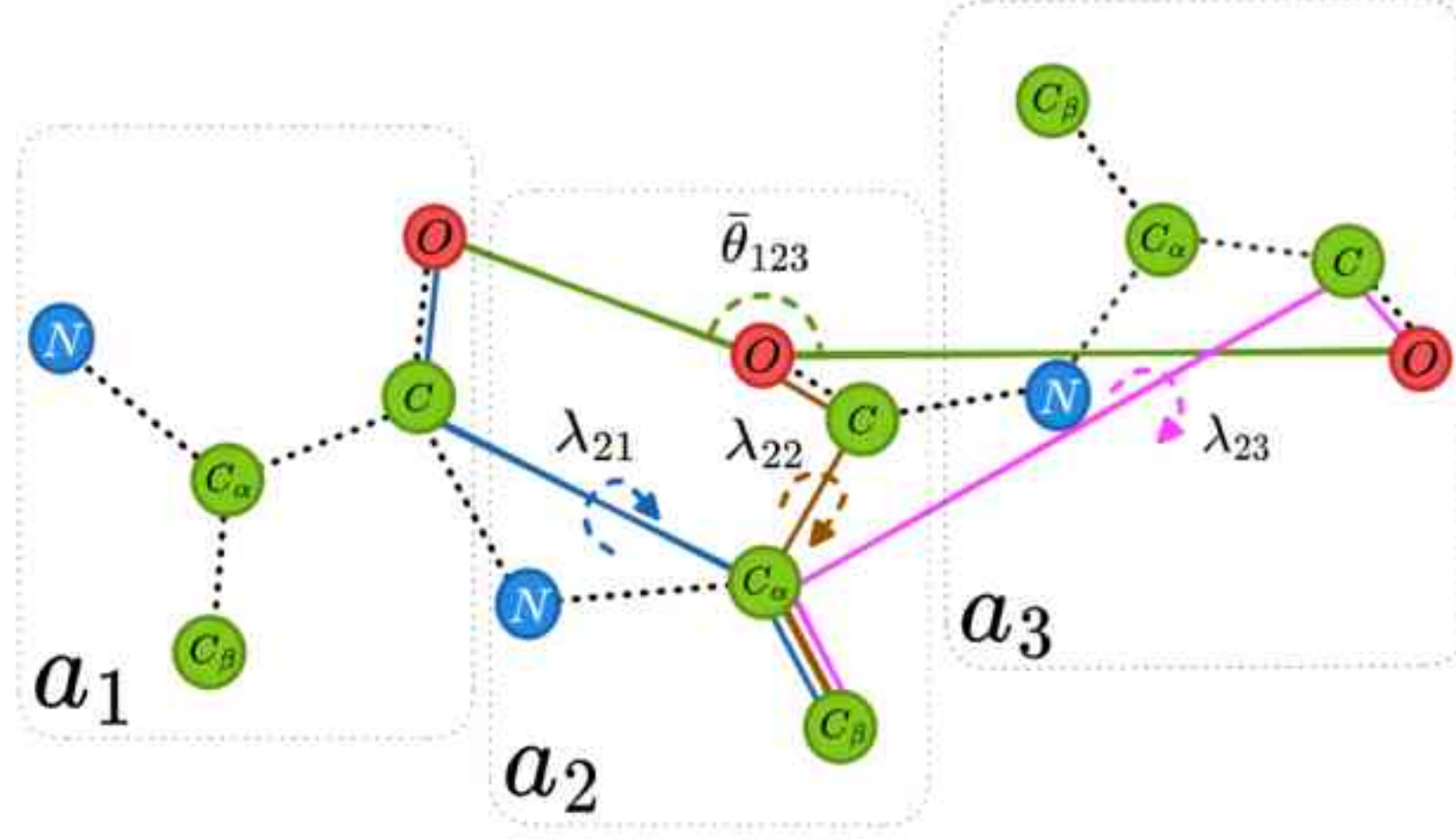
Şekil 4.23 : (ALA, ALA, ALA), (GLU, ALA, LEU), (GLY, ALA, GLY) ve (THR, ALA, THR) amino asit üçlülerinin ayrı ayrı tüm parametre bin kombinasyonlarının ve karşılık gelen $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerjilerin gösterimi.

Görüleceği üzere, ψ - ϕ uzayında her amino asit üçlüsü farklı sayıda parametre bin kombinasyonlarında ve farklı miktarlarda gözlemlenmektedir. Özellikle (GLY, ALA, GLY) amino asit üçlüsü diğerlerinden çok daha yüksek konformasyonel tercih yapabildiğini göstermektedir.

4.2.3.4 Amino Asit Üçlüsünün Merkezindeki Amino Asitin $C_\alpha - C_\beta$ Kovalent Bağının Üçlüdeki Tüm $C = O$ Kovalent Bağlarına Göre Hizalanmasını λ Dihedral ve $\bar{\theta}$ Bağ Açılı ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\lambda-\bar{\theta}}$

Amino asitler üçlülerinde geometrik tercihleri ortaya çıkarmaya çalışırken; sadece $C_\alpha - C_\beta$ kovalent bağlarının ve $C = O$ kovalent bağlarının kendi aralarındaki hizalanmalarını ele almak, diğer geometrik tercihleri gözardı etmemiz anlamına gelecektir. Diğer bir

deyişle, amino asit üçlülerinin bir tarafını oluşturan karbonil grupları ile diğer tarafını oluşturan kenar zincir grupları arasındaki hizalanmanında gözönüne alınması gerektiğini düşünmekteyiz.



Şekil 4.24 : Her (a_1, a_2, a_3) amino asit üçlüsünün merkezindeki a_2 amino asitinin $C_\beta^2 - C_\alpha^2$ kovalent bağının üç amino asitteki $C = O$ kovalent bağına göre hizalanmaları ve geometrik tercihlerini belirlemek için kullanılan λ_{21} , λ_{22} , λ_{23} dihedral açıları ve $\bar{\theta}_{123}$ eğilme açısı terimlerinin gösterimi.

Bu amaçla iki farklı istatistiksel enerji terimi geliştirilmiştir. Amino asit üçlüsünün merkezindeki amino asitin (yani, a_2 'nin) $C_\alpha^2 - C_\beta^2$ kovalent bağının amino asit üçlüsündeki, üç $C = O$ kovalent bağı ile hizalanması betimleyecek $\lambda_{21} \equiv \angle C_\beta^2 C_\alpha^2 C^1 O^1$, $\lambda_{22} \equiv \angle C_\beta^2 C_\alpha^2 C^2 O^2$, $\lambda_{23} \equiv \angle C_\beta^2 C_\alpha^2 C^3 O^3$ dihedral açıları ve $\bar{\theta}_{123} \equiv \angle O^1 O^2 O^3$ bağ açısı tanımlanmıştır. Bu parametrelere bağlı olarak hesaplanacak $E_{\text{üçlü}}^{\lambda-\bar{\theta}}((a_1, a_2, a_3), \lambda_{21}, \lambda_{22}, \lambda_{23}, \bar{\theta}_{123})$ fonksiyonunun matematiksel formu Eşitlik 4.18'de verilmiştir.

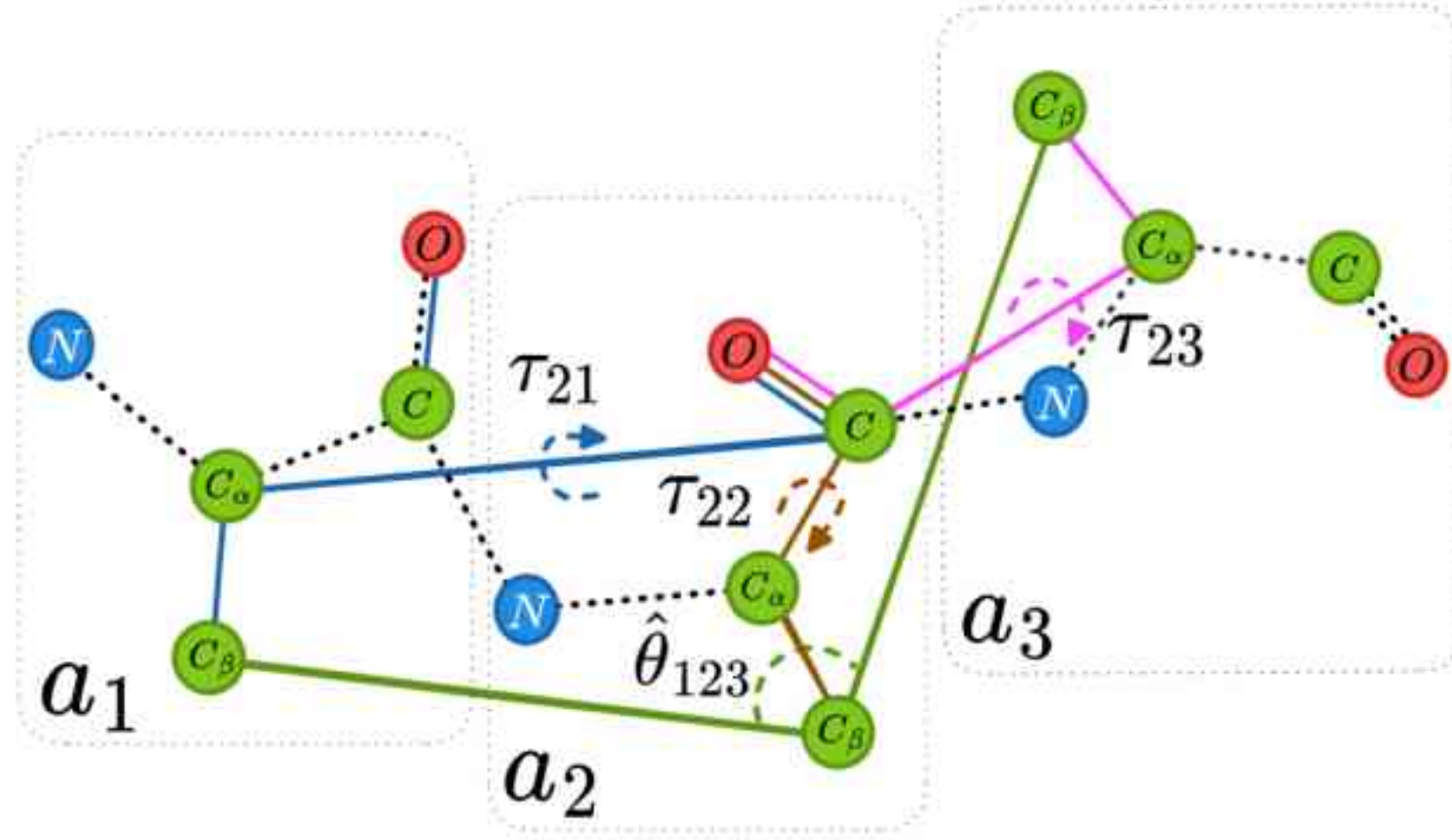
$$E_{\text{üçlü}}^{\lambda-\bar{\theta}}((a_1, a_2, a_3), \lambda_{21}, \lambda_{22}, \lambda_{23}, \bar{\theta}_{123}) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^{\lambda-\bar{\theta}}((a_1, a_2, a_3), \lambda_{21}, \lambda_{22}, \lambda_{23}, \bar{\theta}_{123})}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.18)$$

4.2.3.5 Amino Asit Üçlüsünün Merkezindeki Amino Asitin $C = O$ Kovalent Bağının Üçlüdeki Tüm $C_\alpha - C_\beta$ Kovalent Bağlarına Göre Hizalanmasını τ Dihedral ve $\hat{\theta}$ Bağ Açıları ile Tasvir Eden Enerji Terimi, $E_{\text{üçlü}}^{\tau-\hat{\theta}}$

Amino asit üçlülerindeki $C = O$ ve $C_\alpha - C_\beta$ kovalent bağlarının birbirlerine göre hizalanmalarına odaklanan ve geometrik tercihlerini belirlemek için geliştirdiğimiz ikinci istatistiksel enerji $E_{\text{üçlü}}^{\tau-\hat{\theta}}((a_1, a_2, a_3), \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123})$ terimimiz; amino asit

üçlüsünün merkezindeki amino asitin (yani, a_2 'nin) $C^2 = O^2$ kovalent bağının amino asit üçlüsündeki, üç $C_\alpha - C_\beta$ kovalent bağları arasında tanımlanan $\tau_{21} \equiv \angle O^2 C^2 C_\alpha^1 C_\beta^1$, $\tau_{22} \equiv \angle O^2 C^2 C_\alpha^2 C_\beta^2$, $\tau_{23} \equiv \angle O^2 C^2 C_\alpha^3 C_\beta^3$ dihedral ve $\hat{\theta}_{123} \equiv \angle C_\beta^1 C_\beta^2 C_\beta^3$ bağ açılarının bir fonksiyonudur ve grafiksel gösterimi Şekil 4.25'de verilmiştir. Eşitlik 4.19'da $E_{\text{üçlü}}^{\tau-\hat{\theta}}((a_1, a_2, a_3), \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123})$ matematiksel formu verilmiştir.

$$E_{\text{üçlü}}^{\tau-\hat{\theta}}((a_1, a_2, a_3), \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123}) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^{\tau-\hat{\theta}}((a_1, a_2, a_3), \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123})}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.19)$$

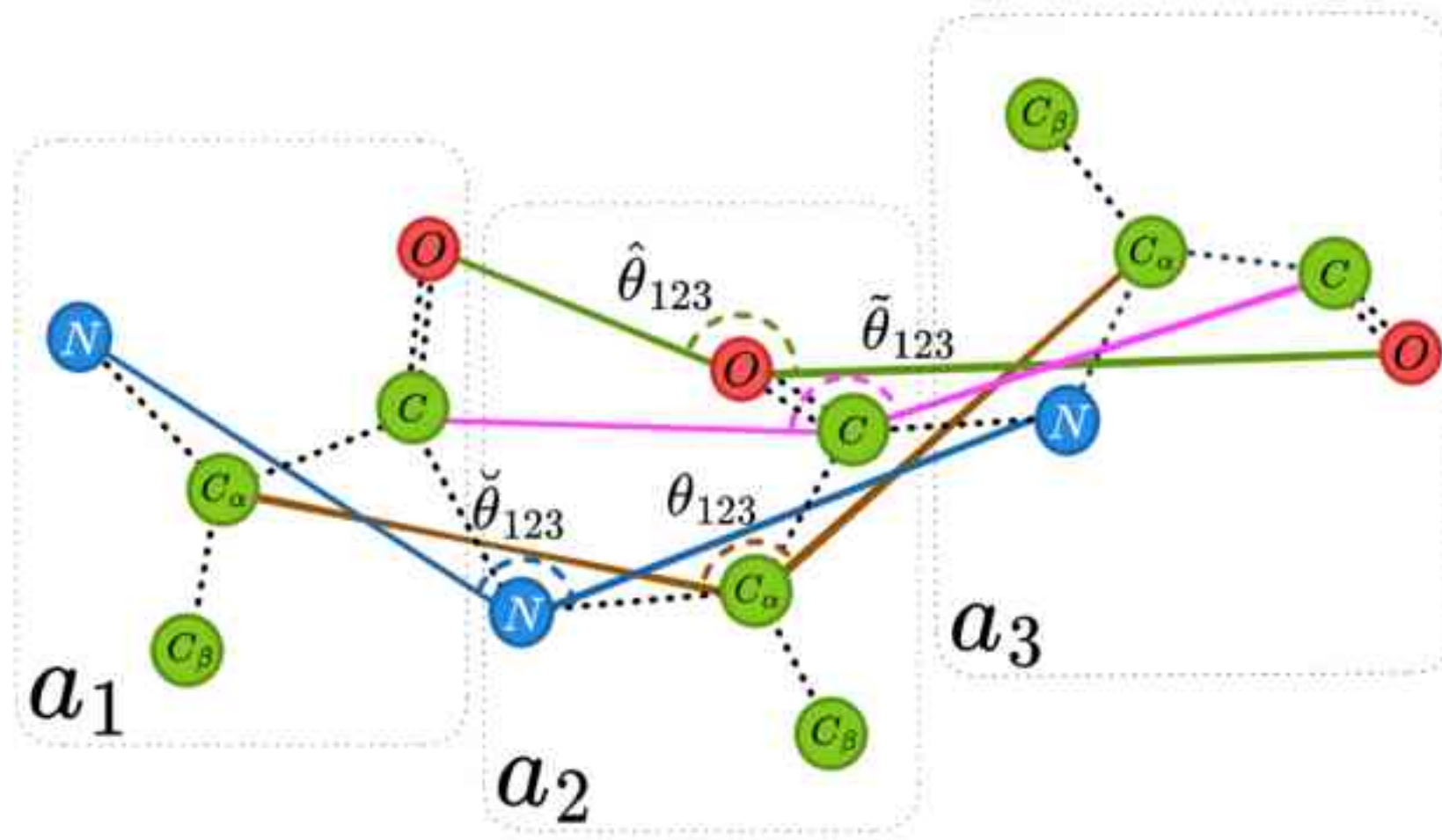


Şekil 4.25 : Her (a_1, a_2, a_3) amino asit üçlüsünün merkezindeki a_2 amino asitinin $O = C$ kovalent bağının üç amino asitteki $C_\alpha - C_\beta$ kovalent bağına göre hizalanmaları ve geometrik tercihlerini belirlemek için kullanılan τ_{21} , τ_{22} , τ_{23} dihedral açıları ve $\hat{\theta}_{123}$ bağ açısı terimlerinin gösterimi.

4.2.3.6 Amino Asit Üçlüsündeki Özdeş Omurga Atomları Arasında Tanımlanan Dört Farklı θ Bağ Açıları ile Tasvir Edilen Enerji Terimi, $E_{\text{üçlü}}^\theta$

Amino asit üçlülerinde özdeş olan atomlar arasında $\check{\theta}_{123} \equiv \angle N^1 N^2 N^3$, $\theta_{123} \equiv \angle C_\alpha^1 C_\alpha^2 C_\alpha^3$, $\tilde{\theta}_{123} \equiv \angle C^1 C^2 C^3$ ve $\bar{\theta}_{123} \equiv \angle O^1 O^2 O^3$ şeklinde tanımlanan ve grafiksel gösterimi Şekil 4.26'da verilen bağ açılarının bir fonksiyonu olan $E_{\text{üçlü}}^\theta((a_1, a_2, a_3), \check{\theta}_{123}, \theta_{123}, \tilde{\theta}_{123}, \bar{\theta}_{123})$ enerji terimi Eşitlik 4.20'de verilmiştir.

$$E_{\text{üçlü}}^\theta((a_1, a_2, a_3), \check{\theta}_{123}, \theta_{123}, \tilde{\theta}_{123}, \bar{\theta}_{123}) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^\theta((a_1, a_2, a_3), \check{\theta}_{123}, \theta_{123}, \tilde{\theta}_{123}, \bar{\theta}_{123})}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.20)$$



Şekil 4.26 : Her (a_1, a_2, a_3) amino asit üçlüsündeki özdeş omurga atomlarının oluşturdukları θ_{123} , $\hat{\theta}_{123}$, $\tilde{\theta}_{123}$ ve $\bar{\theta}_{123}$ bağ açısı terimlerinin gösterimi.

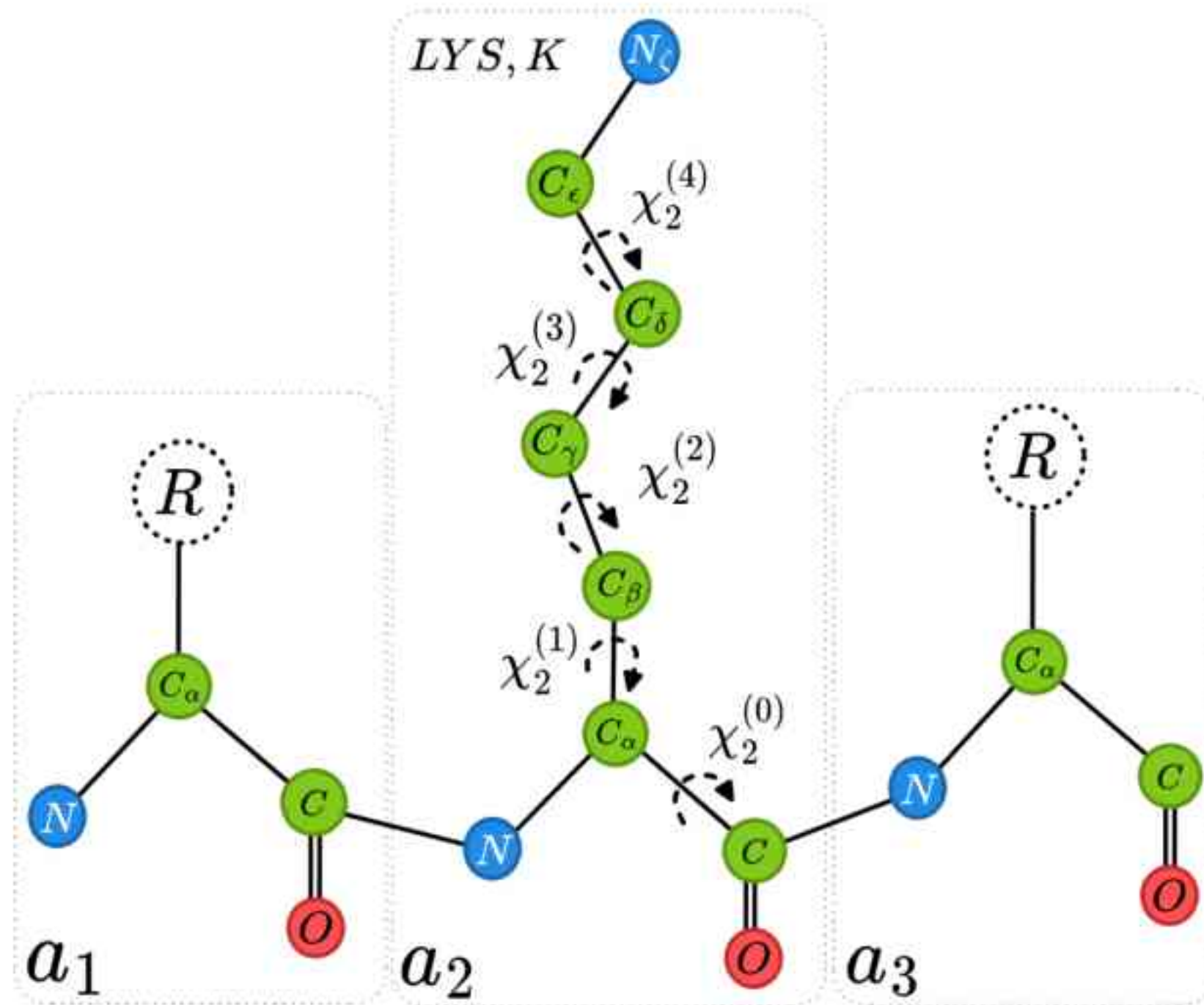
4.2.3.7 Amino Asit Üçlüsünün Merkezindeki Amino Asitin Kenar Zincir Atomları Arasında Tanımlanan χ Dihedral Açılar ile Tasvir Edilen Enerji Terimi, $E_{\text{üçlü}}^{\chi}$

Proteinleri oluşturan amino asitlerin omurga atomları (N, C_{α}, C ve O) aynı olduğu için her amino asitin fiziko-kimyasal doğasını kenar zincir R grupları (bkz. Şekil 4.27) belirlemektedir. Mesela, kenar zincirinde sadece C atomları bulunan amino asitler hidrofobik, yüksek elektronegatifliğe sahip N, O ve S atomları bulunan amino asitler polar ve N^+ veya O^- atomları bulunan amino asitler yüklü amino asitler olarak kabul edilmektedir. Atomların farklı fiziko-kimyasal karakteristikleri proteinlerin katlanma yollarını ve 3-boyutlu katlı yapılarını belirleyen en temel unsurdur. İşte amino asitlerin fiziko-kimyasal özelliklerini belirleyen kenar zincir grubunu oluşturan atomların birbirlerine göre oryantasyonları (Şekil 4.27’de gösterilen χ dihedral açıları) incelendiğinde: χ dihedral açılarının belirli değerlerde öbeklendiği gözlemlenmiştir. Bu nedenle, amino asit kenar zincirlerinin χ dihedral açılara dayalı transfer edilebilir bir istatistiksel enerji terimi $E_{\text{üçlü}}^{\chi}$ geliştirilmiştir. Bu bağlamda, her amino asiti içinde bulunduğu çevreden bağımsız olarak tek başına ele almak yerine, (a_1, a_2, a_3) amino asit üçlüsünün merkezinde bulunan amino asit olarak ele alınıp ayrı ayrı sınıflandırılmıştır (bkz. Şekil 4.27). Diğer bir deyişle, her amino asit türü için $20 \times 20 = 400$ farklı veri seti oluşturularak, χ_2 dihedral açıları daha da spesifik hale dönüştürülmüştür. Böylelikle komşu iki amino asitin (a_1 ve a_3) kenar zincir gruplarının, merkezdeki a_2 amino asitinin kenar zincir χ_2 dihedral açı değerleri dağılımına katkısı göz önüne alınmış olmaktadır. Belirtilmesi gerekir ki; amino asitlerin kenar zincir gruplarının büyüklüğüne bağlı olarak χ_2 dihedral açı terimi sayısı değişmektedir. Şekil 4.27’te en yüksek (dört tane) χ_2 dihedral

açılarına sahip amino asitlerden lizin (LYS, K) örnek olarak seçilmiştir. $\chi_2^{(0)}$ dihedral açısı esasen kenar zincir dihedral açısı olmamasına rağmen, kenar zincir atomlarının konformasyonel tercihleri ile omurga karbonil C ve O atomları ile ilişkilendirilerek geometri spesifikliği arttırılmaya çalışılmıştır. $E_{\text{üçlü}}^{\chi}$ enerji terimi aşağıda verilen eşitlik ile hesaplanmaktadır.

$$E_{\text{üçlü}}^{\chi} \left((a_1, a_2, a_3), \chi_2^{(0)}, \chi_2^{(1)}, \chi_2^{(2)}, \chi_2^{(3)}, \chi_2^{(4)} \right) = -kT \ln \left[1 + \frac{N_{\text{üçlü}}^{\chi} \left((a_1, a_2, a_3), \chi_2^{(0)}, \chi_2^{(1)}, \chi_2^{(2)}, \chi_2^{(3)}, \chi_2^{(4)} \right)}{\rho_{\text{üçlü}}(a_1, a_2, a_3)} \right] \quad (4.21)$$

Eşitlikte $\chi_2^{(0)} \equiv \angle O^2 C^2 C_{\alpha}^2 C_{\beta}^2$, $\chi_2^{(1)} \equiv \angle N^2 C_{\alpha}^2 C_{\beta}^2 C_{\gamma}^2$, $\chi_2^{(2)} \equiv \angle C_{\alpha}^2 C_{\beta}^2 C_{\gamma}^2 C_{\delta}^2$, $\chi_2^{(3)} \equiv \angle C_{\beta}^2 C_{\gamma}^2 C_{\delta}^2 C_{\epsilon}^2$ ve $\chi_2^{(4)} \equiv \angle C_{\gamma}^2 C_{\delta}^2 C_{\epsilon}^2 N_{\zeta}^2$ lizin amino asiti kenar zincir atomları için tanımlanan dihedral açı terimlerini ve $N_{\text{üçlü}}^{\chi} \left((a_1, a_2, a_3), \chi_2^{(0)}, \chi_2^{(1)}, \chi_2^{(2)}, \chi_2^{(3)}, \chi_2^{(4)} \right)$ terimi ise bu açı değerlerinde gözlemlenen toplam (a_1, a_2, a_3) amino asit üçlüsü sayısını temsil etmektedir.

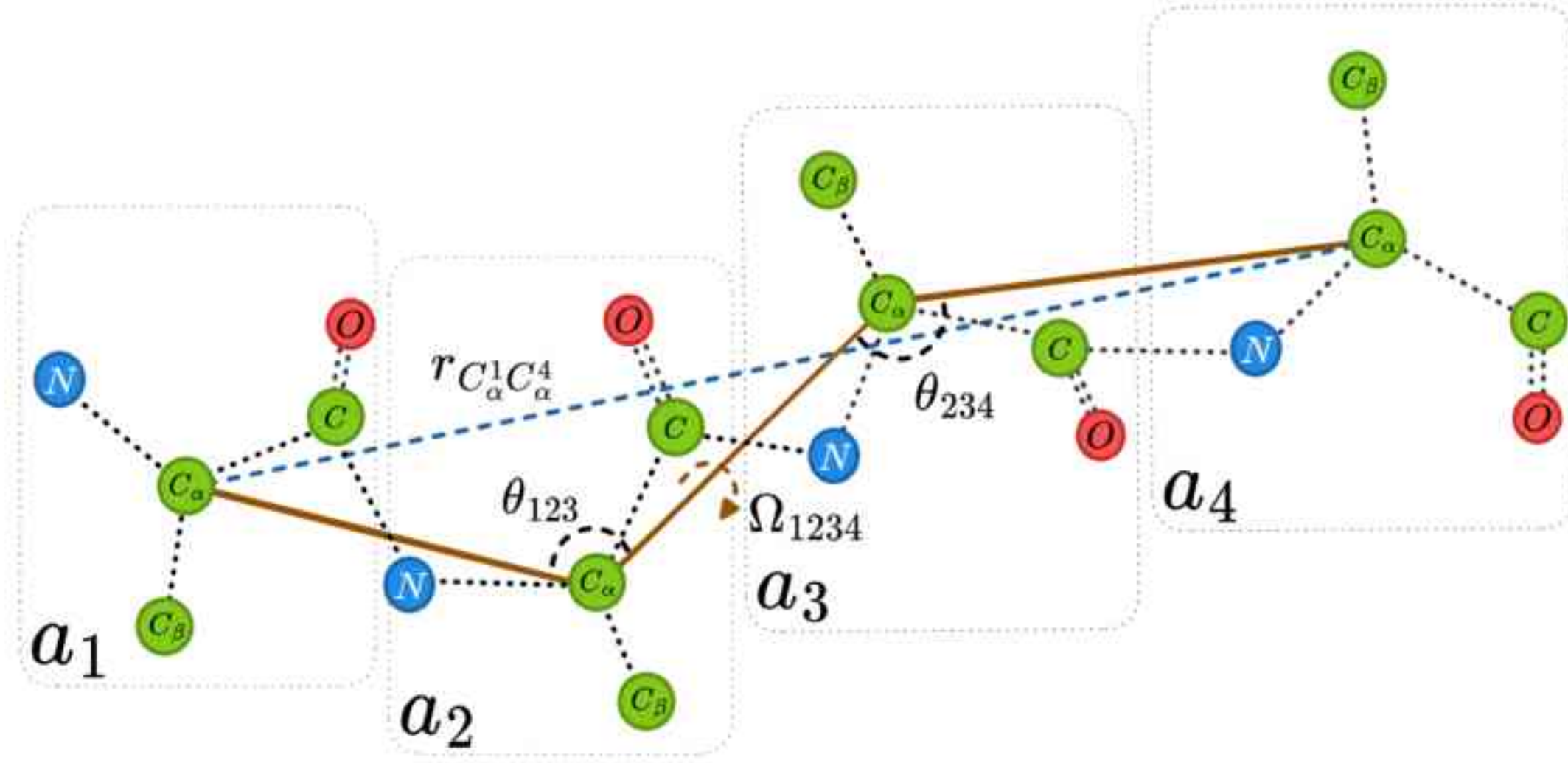


Şekil 4.27 : (a_1, a_2, a_3) amino asit üçlüsü için pozitif yüklü lizin amino asitinin merkezde olduğu yani $a_2 \equiv \text{LYS}$ seçildiği durumda; χ_2 dihedral açılarının ve bu açılar hangi atom dörtlüleri gözönüne alındığının gösterimi.

4.2.3.8 Amino Asit Dörtlüleri için Etkileşim Enerji Terimleri, $E_{\text{dörtlü}}^{C_{\alpha}}$ ve $E_{\text{dörtlü}}^{NCOC_{\beta}}$

Protein modellemelerinde, tamamıyla sarmallardan oluşan (mesela, 3-sarmal veya 4-sarmal demetli) katlı yapıya sahip proteinlerde en sık karşılaşılan sorunlardan bir tanesi; dinamik konformasyonel dönüşümler sırasında düşük enerjili ve katlı yapıyı çağrıştıran

fakat gerçek katlı yapıdan farklı tuzak konformasyonlarda hapsolunabilmektedir. Bu durumun başlıca sebeplerinden birisi sarmallar doğru oluştursa da, bu sarmalların birbirine göre pozisyonlarında ayna görüntüsü (mirror image) oluşabilmektedir.



Şekil 4.28 : Protein zinciri üzerinde ardışık amino asit dörtlülerinin C_α atomlarının geometrik tercihlerini belirlemek için kullanılan $r_{C_\alpha^1 C_\alpha^4}$, mesafesi, $\theta_{123}^{C_\alpha}$ ve $\theta_{234}^{C_\alpha}$ bağ açısı ve $\Omega_{1234}^{C_\alpha}$ dihedral açısı terimlerinin gösterimi.

Bu tarz ayna görüntüsü oluşumunu ve başka tuzaklanma mekanizmalarını engelleyebilmek için, protein zinciri üzerinde ardışık olan dört amino asitin sadece C_α atomlarının koordinatlarına göre tanımlanan ve Şekil.4.28’de grafiksel gösterimi yapılan; bir tane dihedral açısı $\Omega_{1234}^{C_\alpha} \equiv \angle C_\alpha^1 C_\alpha^2 C_\alpha^3 C_\alpha^4$, iki tane bağ açısı $\theta_{123}^{C_\alpha} \equiv \angle C_\alpha^1 C_\alpha^2 C_\alpha^3$ ile $\theta_{234}^{C_\alpha} \equiv \angle C_\alpha^2 C_\alpha^3 C_\alpha^4$ ve $r_{C_\alpha^1 C_\alpha^4}$ öklid mesafe değişkenlerinin ayrık aralıklarına bağlı $E_{\text{dörtlü}}^{C_\alpha} \left((a_1, a_2, a_3, a_4), r_{C_\alpha^1 C_\alpha^4}, \theta_{123}^{C_\alpha}, \theta_{234}^{C_\alpha}, \Omega_{1234}^{C_\alpha} \right)$ enerji fonksiyonunun matematiksel formu Eşitlik 4.22 verilmiştir.

$$E_{\text{dörtlü}}^{C_\alpha} \left((a_1, a_2, a_3, a_4), r_{C_\alpha^1 C_\alpha^4}, \theta_{123}^{C_\alpha}, \theta_{234}^{C_\alpha}, \Omega_{1234}^{C_\alpha} \right) = -kT \ln \left[1 + \frac{N_{\text{dörtlü}}^{C_\alpha} \left((a_1, a_2, a_3, a_4), r_{C_\alpha^1 C_\alpha^4}, \theta_{123}^{C_\alpha}, \theta_{234}^{C_\alpha}, \Omega_{1234}^{C_\alpha} \right)}{\rho_{\text{dörtlü}}(a_1, a_2, a_3, a_4)} \right] \quad (4.22)$$

Eşitlik.4.22’deki pay ve payda terimlerinin hepsi amino asit üçlüsü için yapılan tanımlamalar ile birebir aynı olup sadece amino üçlüsü yerine dörtlüsü gözönüne alınmıştır.

Diğer taraftan, amino asit üçlülerinde özdeş atomlar arasında tanımladığımız bağ açıları yerine; amino asit dörtlüleri için yine özdeş atomlar arasında $\Omega_{1234}^N \equiv \angle N^1 N^2 N^3 N^4$, $\Omega_{1234}^C \equiv \angle C^1 C^2 C^3 C^4$, $\Omega_{1234}^O \equiv \angle O^1 O^2 O^3 O^4$ ve $\Omega_{1234}^{C_\beta} \equiv \angle C_\beta^1 C_\beta^2 C_\beta^3 C_\beta^4$

şeklinde tanımlanan dihedral açılarının fonksiyonu olan ve matematiksel formu Eşitlik 4.23'te verilen $E_{\text{dörtlü}}^{NCOC\beta} \left((a_1, a_2, a_3, a_4), \Omega_{1234}^N, \Omega_{1234}^C, \Omega_{1234}^O, \Omega_{1234}^{C\beta} \right)$ istatistiksel enerji terimi geliştirilmiştir.

$$E_{\text{dörtlü}}^{NCOC\beta} \left((a_1, a_2, a_3, a_4), \Omega_{1234}^N, \Omega_{1234}^C, \Omega_{1234}^O, \Omega_{1234}^{C\beta} \right) = -kT \ln \left[1 + \frac{N_{\text{dörtlü}}^{NCOC\beta} \left((a_1, a_2, a_3, a_4), \Omega_{1234}^N, \Omega_{1234}^C, \Omega_{1234}^O, \Omega_{1234}^{C\beta} \right)}{\rho_{\text{dörtlü}}(a_1, a_2, a_3, a_4)} \right] \quad (4.23)$$

Amino asit dörtlüsü enerji terimleri $E_{\text{dörtlü}}^{C\alpha}$ ve $E_{\text{dörtlü}}^{NCOC\beta}$ için referans durumu tanımlayan $\rho_{\text{dörtlü}}(a_1, a_2, a_3, a_4)$ mol kesri, amino asit üçlülerinin mol kesri $\rho_{\text{üçlü}}(a_1, a_2, a_3)$ 'ye benzer şekilde aşağıda verilen Eşitlik 4.24 ile hesaplanmaktadır.

$$\rho_{\text{dörtlü}}(a_1, a_2, a_3, a_4) = \frac{N_{\text{dörtlü}}(a_1, a_2, a_3, a_4)}{\sum_{i=1}^{20} \sum_{j=1}^{20} \sum_{k=1}^{20} \sum_{l=1}^{20} N_{\text{dörtlü}}(a_1^i, a_2^j, a_3^k, a_4^l)} \quad (4.24)$$

4.2.4 Protein Konformasyonu Toplam Enerjisi

Şu ana kadar geliştirdiğimiz transfer edilebilir enerji terimleri kullanılarak bir protein konformasyonunun toplam enerjisi E_{konf} aşağıda verilen Eşitlik 4.25 ile hesaplanmaktadır.

Eşitlik 4.25'teki $\tilde{E}_{\text{çift}}$ enerji terimi protein zinciri üzerinde birbirine üç kovalent bağı uzaklığından daha yakın olan ve/veya aynı amino asitin içinde bulunan atom çiftleri hariç, mümkün olan diğer tüm atom çiftleri üzerinden hesaplanan toplam atom çifti etkileşme enerjisini ifade etmektedir. Atom çiftleri, zincir üzerindeki mesafeleri d 'ye bağlı olarak ayrı ayrı hesaplanmaktadır. Bu nedenle, atom çifti toplam enerjisi $\tilde{E}_{\text{çift}}$, Eşitlik 4.25'te iki ayrı toplam (birinci toplam zincir mesafesi d ve ikinci toplam i ve $j > i + 1$ atom çiftleri üzerinden) ile formüle edilmiştir.

Diğer taraftan, $\tilde{E}_{\text{Hb}}^{N\cdots O}$ enerji terimi ise $N - H \cdots O$ hidrojen bağı toplam enerjilerini temsil etmektedir. Bu enerji terimlerindeki toplamlar ise amino asit çiftleri üzerinden alınmaktadır. Benzer şekilde, amino asit üçlülerini için geliştirilen tüm $\tilde{E}_{\text{üçlü}}$ enerji terimlerindeki toplamlar, proteindeki amino asit üçlülerini (toplam $N - 2$ tane) ve amino asit dörtlülerini için geliştirilen tüm $\tilde{E}_{\text{dörtlü}}$ enerji terimindeki toplam ise proteindeki amino asit dörtlülerini (toplam $N - 3$ tane) üzerinden alınmaktadır.

$$\begin{aligned}
E_{\text{konf}} = & \tilde{E}_{\text{çift}} + \omega_{\text{Hb}} \tilde{E}_{\text{Hb}}^{N\cdots O} + \omega_{\text{dörtlü}}^{C_\alpha} \tilde{E}_{\text{dörtlü}}^{C_\alpha} + \omega_{\text{dörtlü}}^{NCOC_\beta} \tilde{E}_{\text{dörtlü}}^{NCOC_\beta} + \omega_{t_1} \tilde{E}_{\text{üçlü}}^{\eta-\theta} \\
& + \omega_{t_2} \tilde{E}_{\text{üçlü}}^{\xi-\tilde{\theta}} + \omega_{t_3} \tilde{E}_{\text{üçlü}}^{\psi-\phi} + \omega_{t_4} \tilde{E}_{\text{üçlü}}^{\lambda-\bar{\theta}} + \omega_{t_5} \tilde{E}_{\text{üçlü}}^{\tau-\hat{\theta}} + \omega_{t_6} \tilde{E}_{\text{üçlü}}^{\theta} + \omega_{t_7} \tilde{E}_{\text{üçlü}}^{\chi} \\
= & \sum_{d=1}^7 \sum_{\substack{i=1 \\ j>i+1}}^{N_{\text{atom}}} \omega_{\text{çift}}(a\tilde{a}_i, a\tilde{a}_j) E_{\text{çift}}^d(\tilde{a}_i, \tilde{a}_j, r_{ij}) \\
& + \omega_{\text{Hb}} \sum_{\substack{i=1 \\ j>i+2}}^N E_{\text{Hb}}^{N\cdots O}((aN)_i, (aO)_j, r_1, r_2, r_3, r_4) \\
& + \omega_{\text{dörtlü}}^{C_\alpha} \sum_{i=1}^{N-3} E_{\text{dörtlü}}^{C_\alpha}((a_1, a_2, a_3, a_4)_i, r_{C_\alpha^1 C_\alpha^4}, \theta_{123}^{C_\alpha}, \theta_{123}^{C_\alpha}, \Omega_{1234}^{C_\alpha}) \\
& + \omega_{\text{dörtlü}}^{NCOC_\beta} \sum_{i=1}^{N-3} E_{\text{dörtlü}}^{NCOC_\beta}((a_1, a_2, a_3, a_4)_i, \Omega_{1234}^N, \Omega_{1234}^C, \Omega_{1234}^O, \Omega_{1234}^{C_\beta}) \\
& + \omega_{t_1} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\eta-\theta}((a_1, a_2, a_3)_i, \eta_{21}, \eta_{23}, \eta_{13}, \theta_{123}) \\
& + \omega_{t_2} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\xi-\tilde{\theta}}((a_1, a_2, a_3)_i, \xi_{12}, \xi_{23}, \xi_{13}, \tilde{\theta}_{123}) \\
& + \omega_{t_3} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3)_i, \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23}) \\
& + \omega_{t_4} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\lambda-\bar{\theta}}((a_1, a_2, a_3)_i, \lambda_{21}, \lambda_{22}, \lambda_{23}, \bar{\theta}_{123}) \\
& + \omega_{t_5} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\tau-\hat{\theta}}((a_1, a_2, a_3)_i, \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123}) \\
& + \omega_{t_6} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\theta}((a_1, a_2, a_3)_i, \theta_{123}, \check{\theta}_{123}, \tilde{\theta}_{123}, \bar{\theta}_{123}, \hat{\theta}_{123}) \\
& + \omega_{t_7} \sum_{i=1}^{N-2} E_{\text{üçlü}}^{\chi}((a_1, a_2, a_3)_i, \chi_2^{(0)}, \chi_2^{(1)}, \chi_2^{(2)}, \chi_2^{(3)}, \chi_2^{(4)}) \tag{4.25}
\end{aligned}$$

Yine, Eşitlik 4.25'deki tüm ω parametreleri enerji ölçekleme parametreleridir. $\omega_{\text{çift}}(a\tilde{a}_i, a\tilde{a}_j)$ terimi hariç diğerleri birer sabit olarak tanımlanmışlardır. $\omega_{\text{çift}}(a\tilde{a}_i, a\tilde{a}_j)$ terimi ise dinamik bir ölçekleme parametresidir. Dinamik konformasyonel geçişler süresince, ikincil yapılar oluşmadan gerçekleşecek hidrofobik çökme veya kenar zincir atom sayısı yüksek olan (mesela, PHE, TYR, TRP vd.) amino asitlerin bir araya toplanıp düşük enerjili tuzak konformasyon oluşturdıkları durumlarda; kenar zincir atomları arasındaki etkileşimleri ölçekleyerek bu tür konformasyonların ziyaret edilmesinin önüne geçecek şekilde sürekli değişebilmektedir.

Bu dinamik ölçekleme parametresi birbirleriyle etkileşen iki atomun ait olduğu amino asitlerin her biri için Eşitlik 4.26 kullanılarak hesaplanan $\omega(a_i)$ değerlerine bağlıdır.

$$\omega(a_i) = N_{\text{atom}}^{a_{i-1}} \sum_{k=1}^7 \delta_{t_k}^{a_{i-1}} \omega_{t_k} + N_{\text{atom}}^{a_i} \sum_{k=1}^7 \delta_{t_k}^{a_i} \omega_{t_k} + N_{\text{atom}}^{a_{i+1}} \sum_{k=1}^7 \delta_{t_k}^{a_{i+1}} \omega_{t_k} \quad (4.26)$$

Bu eşitlikteki ω_{t_k} amino asit üçlülerinin ağırlığını, $N_{\text{atom}}^{a_i}$ amino asit üçlüsündeki i 'nci amino asitin sahip olduğu toplam atom sayısını temsil etmektedir. $\delta_{t_k}^{a_i}$ terimi ise Eşitlik 4.27'de gösterildiği amino asit üçlüsü enerjisinin aldığı değere göre sadece 0 ila 1 değeri alan bir delta fonksiyonudur.

$$\delta_{t_k}^{a_i} = \begin{cases} 1, & \text{eğer } E_{t_k}^{a_i} < 0 \\ 0, & \text{eğer } E_{t_k}^{a_i} = 0 \end{cases} \quad (4.27)$$

Etkileşen amino asitler için hesaplanan $\omega(a_i)$ değerlerinin toplamı Eşitlik 4.28'de gösterildiği gibi normalize edilerek, $\omega_{\text{çift}}(a_i, a_j)$ değerleri hesaplanmaktadır.

$$\omega_{\text{çift}}(a_i, a_j) = \frac{\omega(a_i) + \omega(a_j)}{(N_{\text{atom}}^{a_{i-1}} + N_{\text{atom}}^{a_i} + N_{\text{atom}}^{a_{i+1}} + N_{\text{atom}}^{a_{j-1}} + N_{\text{atom}}^{a_j} + N_{\text{atom}}^{a_{j+1}}) \sum_{k=1}^7 \omega_{t_k}} \quad (4.28)$$

4.2.5 Monte Carlo Protein Dinamiği Algoritması

Monte Carlo Simülasyon yöntemi, deterministik moleküler dinamik simülasyon yönteminin aksine stokastik bir yöntem olup, dinamik süresince gerçekleşen tüm olaylar veya değişimler gelişigüzel (rastlantısal) gerçekleşmektedir [91]. Mesela, dinamik süresince bir proteinde hangi amino asit veya atomların nasıl hareket ettirileceği, hareket sonrası elde edilen yeni konformasyona geçişin kabul edilip edilmemesi gibi tüm süreçler gelişigüzel bir şekilde gerçekleşmektedir.

Monte Carlo protein modelimizde dinamiği belirleyen en temel unsurlardan hareket setleri; protein zincirinde omurga atomları arasında tanımlanan ψ ve ϕ dihedral açılarının (bkz. Şekil 4.22) ve kenar zincir atomları arasında tanımlanan χ dihedral açılarının (bkz. Şekil 4.27) değiştirilmesine (döndürülmesine) dayanmaktadır. Bununla birlikte, global ve yerel hareket seti olmak üzere iki ayrı hareket seti kullanılmıştır. Bunlar; (i) ψ ve ϕ dihedral açılarının değiştirilmesi ile protein zinciri üzerinde gerçekleştirilen global rijit dönme hareket seti [92] [93], ve (ii) protein zincir üzerinde ardışık amino asitlerin oluşturduğu bir zincir kesitindeki omurga atomlarının, geometrik olarak kesitten önceki

ve sonraki kısımlarla mümkün olan uyumlu omurga konformasyonlardan birisine geçişi tanımlayan yerel hareket setidir. Yerel harekette sadece yerel zincir kesitindeki atomlar hareket ederken, zincir kesiti dışındaki atomlar sabit kalmaktadır [94].

Monte Carlo protein dinamiği algoritmamız, her zaman adımında tamamı gelişigüzel gerçekleştirilen sırasıyla şu süreçlerden oluşmaktadır: (a) amino asit seçimi yapılması, (b) teşebbüs edilecek hareket türünün (global veya yerel hareket) seçilmesi, (c) seçilen hareket türünde dönme açısı değerinin (global hareket için ortalaması 0 ve standart sapması 2 derece olan, yerel hareket için ortalaması 0 standart sapması 30 derece olan ve kenar zincir atomlarının hareketi için ortalaması 0 ve standart sapması 10 derece olan gaussian dağılımdan [95] elde edilen gelişigüzel dönme açısı değerlerinin) belirlenmesi, (d) teşebbüs edilen hareket sonrası tüm atom çiftleri arasındaki mesafelerin atom çiftlerinin Van der Waals yarıçapları toplamının 0.75 katından daha düşük $\left(0.75 \times (r_{VdW}^i + r_{VdW}^j)\right)$ olup olmadığının test edilmesi, eğer herhangi bir atom çifti arasındaki mesafe bu mesafeden küçük (yani, sterik çarpışma durumu) ise yeni konformasyona geçiş gerçekleştirilmeksizin başa dönülmesi, (d) eğer sterik çarpışma söz konusu değilse konformasyon enerjisi hesaplanarak ve Metropolis kriteri veya algoritması [96] kullanılarak yeni konformasyona geçişin kabul edilip edilmemesinin belirlenmesi, (e) yeni zaman adımına geçilmesi.

Belirtilmesi gerekir ki; her Monte Carlo zaman adımında bir global veya yerel dönme hareketi gerçekleştirilmesine rağmen, 10 tane kenar zincir hareketi uygulanmaktadır. Yani, global ve yerel hareket teşebbüsünün kabul edilip edilmemesinden bağımsız olarak, 10 defa gelişigüzel amino asit seçimi yapılmakta ve seçilen her amino asitin tüm kenar zincir χ dihedral açıları değiştirilerek ayrı ayrı kenar zincir hareketi yapılmaktadır.

Gerçekleştirilen herhangi bir zincir hareketinden sonra sterik çarpışma durumu söz konusu değilse; eski ve yeni konformasyonun enerji farkı $\Delta E_{\text{konf}} = E_{\text{konf}}^{\text{yeni}} - E_{\text{konf}}^{\text{eski}}$ ilgili konformasyon geçişinin kabul edilip edilmemesinde belirleyici rol oynamaktadır. Eğer $\Delta E_{\text{konf}} < 0$ ise yeni konformasyona geçiş direk gerçekleşmektedir. Ters durumda yani $\Delta E_{\text{konf}} > 0$ ise $[0, 1]$ aralığında gelişigüzel sayı üreten fonksiyondan bir v_{rand} sayısı çağırılmakta ve matematiksel formu Eşitlik 4.29'ta verilen Metropolis kriterine [96] göre

$$\begin{cases} \text{kabul,} & \text{eğer } v_{\text{rand}} < \exp\left(-\frac{\Delta E_{\text{konf}}}{k_B T}\right) \\ \text{red,} & \text{eğer } v_{\text{rand}} > \exp\left(-\frac{\Delta E_{\text{konf}}}{k_B T}\right) \end{cases} \quad (4.29)$$

yeni konformasyona geçiş kabul edilmekte veya reddedilmektedir.

4.2.6 Monte Carlo Simülasyonundan Elde Edilen Bulgular

Sadece geliştirdiğimiz transfer edilebilir istatistiksel enerji terimleri içeren ve tüm atomların temsil edildiği Monte Carlo simülasyon yazılımımız kullanılarak; proteinlerin uzun zaman ölçekli davranışları, elde edilen en düşük enerjili ve en düşük RMSD (Kök-Ortalama-Kare-Hatası)'sine sahip konformasyonları incelemekteyiz.

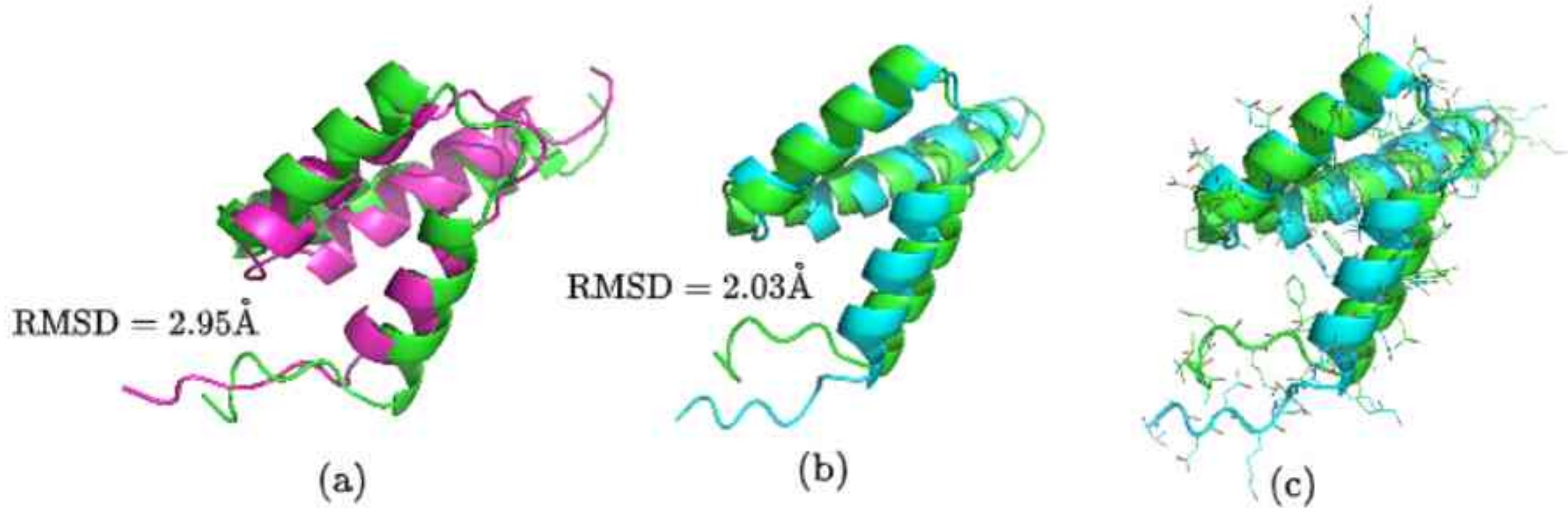
Monte Carlo simülasyon çalışmalarımızda ilk adım olarak protein katlı yapısının çok yüksek sıcaklıkta açılmasını sağlıyoruz. Daha sonra simülasyondan elde edilen ve hiç bir ikincil yapı içermeyen, tamamıyla açık protein konformasyonundan başlayarak farklı sıcaklıklarda katlanma ve açılma geçişlerinin sıklıkla gerçekleşebileceği sıcaklık aralığı taraması yapılmaktadır.

11 tane farklı enerji terimi ve ölçekleme parametresi içeren bir protein konformasyon enerji fonksiyonunun tanımladığı parametre uzayının; hangi bölgelerinde konformasyon uzayının daha efektif taranacağını ortaya çıkarmak bile kendi başına ayrı problem teşkil etmektedir.

Simülasyonlarımızda ilk olarak protein veri bankası kodu 1bdd.pdb, katlı yapısı NMR deneysel yöntemiyle elde edilmiş olan ve 60 amino asit içeren Staphylococcus Aureus Protein A, immünglobulin-bağlanma B bölgesi proteini çalışılmıştır.

1bdd proteini için atom çifti enerji terimi dışındaki tüm enerji terimlerinin ölçeklenme parametreleri birbiriyle aynı ve 2'ye eşit, simülasyon sıcaklığı $T_{sim} = 2.6$ olarak aldığımızda elde edilen en düşük RMSD değerine sahip yapılar Şekil 4.29'da verilmiştir.

Şekil 4.29'da (i) macenta renkli olan deneylerden elde edilen (Protein Veri Bankasından alınan) yapıyı, (ii) yeşil renkli olan simülasyonda açık konformasyondan başlayarak ulaşılan katlı yapıyı ve (iii) mavi renkli olan ise çok düşük sıcaklıkta simülasyona deneyden elde edilen macenta renkli katlı yapıdan başlanarak ulaşılan en düşük enerjili yapıları vermektedir. Mavi renkli yapıları elde etmeye çalışmamızın temel sebebi: protein veri bankasından alınan protein yapılarında su molekülleri bulunmadığı için, özellikle protein katlı yapısının yüzeyindeki atomların su molekülleri ile hangi geometride nasıl etkileştikleri veya hidrojen bağı oluşturdıkları tam bilinmemektedir. Bu nedenle, simülasyon çalışmalarında deneyden elde edilen orijinal katlı yapılar, düşük sıcaklıklarda kullanılan kuvvet alanlarına uygun olacak şekilde dengeye gelmesi veya literatürdeki tabiriyle gevşetilmiş-katlı yapıya (relaxed-native structure) ulaşması beklenir. Biz de

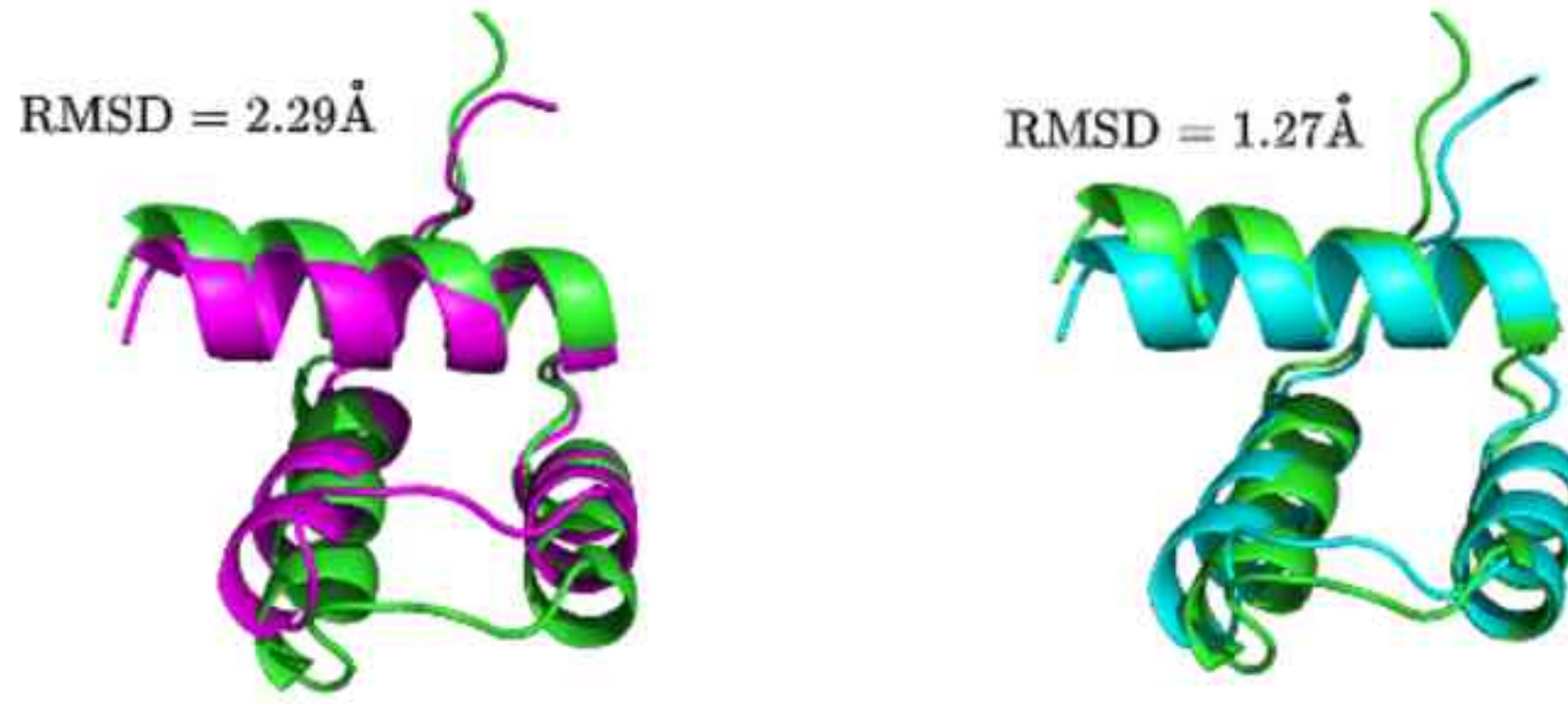


Şekil 4.29 : (a) NMR deneysel yöntemiyle elde edilen 1bdd proteinin orijinal katlı yapısı (macenta renkli) ile simülasyondan elde edilen ve deneyden elde edilen yapıya göre en düşük RMSD değerine sahip protein yapısının (yeşil renkli) hizalanmış hallerinin, (b) gevşetilmiş-katlı yapı (mavi renkli) ve simülasyondan elde edilen katlı yapı (yeşil renkli) hizalanmış hallerinin, (c) (b)'deki yapıların kenar zincir atomları ile birlikte gösterimi.

benzer şekilde deneylerden elde edilen orijinal katlı yapıları, geliştirdiğimiz enerji fonksiyonunu kullanarak en düşük enerjili gevşetilmiş-katlı yapıları elde etmek için kullanmaktayız.

Bu tez kapsamında çalıştığımız diğer protein ise; protein veri bankası kodu 1enh.pdb, katlı yapısı X-ışını kristalografi yöntemiyle elde edilmiş ve 54 amino asiti olan Homeodomain Protein (DNA-bağlanma proteini)'dir. 1enh proteini için 1bdd proteininde kullanılan ölçeklenme parametre değerlerinin aynıları kullanılmıştır. Şekil 4.30'da 1enh proteini için elde edilen en düşük RMSD değerine sahip protein yapısı verilmiştir. Şekil 4.30'da kullanılan renk tanımlamaları, Şekil 4.29'da 1bdd proteini için yapılan tanımlamalar ile birebir aynıdır.

Şekil 4.29 ve 4.30'da görüleceği üzere hala geliştirilme aşamasında olan enerji fonksiyonumuz ve tüm atom temsilli Monte Carlo protein modelimiz, tamamıyla α -sarmallarından oluşan proteinlerde oldukça başarılı sonuçlar vermektedir. Simülasyondan elde edilen yapıların gevşetilmiş-katlı yapıya göre; 1bdd proteini için $\text{RMSD} = 2.03\text{\AA}$ ve 1enh proteini için $\text{RMSD} = 1.27\text{\AA}$ 'lık değerlerine sahip katlı yapıların elde edilmesi; geliştirilen enerji terimlerinin hem protein topolojisini ayırt etmekte hem de ikincil yapıların doğru oluşmasında istenilen düzeydedir.

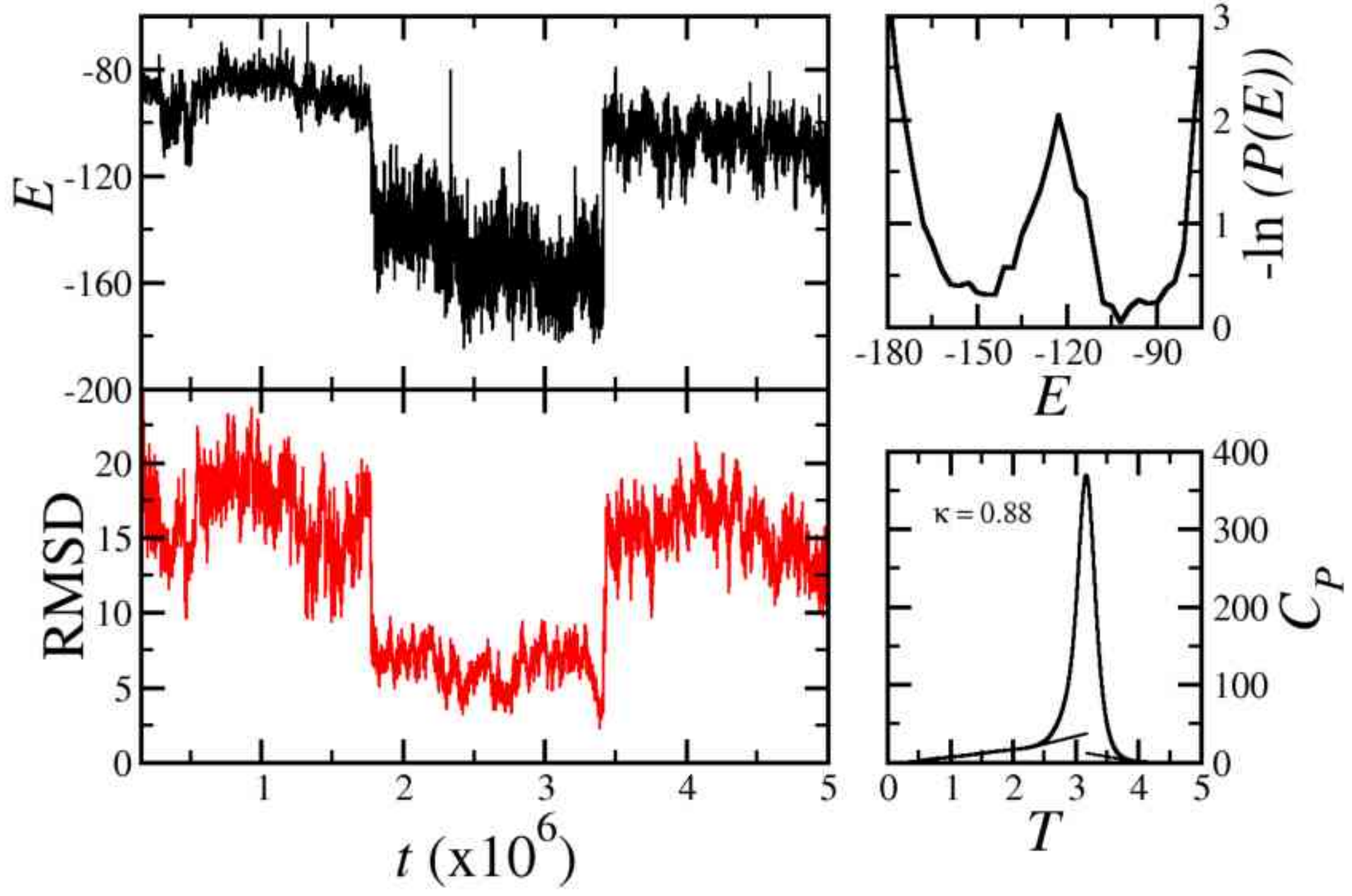


Şekil 4.30 : (a) X-ışını kristalografi deneysel yöntemiyle elde edilen IenH proteinin orijinal katlı yapısı (macenta renkli yapı) ile simülasyondan elde edilen ve deneyden elde edilen yapıya göre en düşük RMSD değerine sahip protein yapısının (yeşil renkli yapı) hizalanmış halinin, (b) gevşetilmiş-katlı yapı (mavi renkli) ve simülasyondan elde edilen katlı yapısının (yeşil renkli) hizalanmış hallerinin gösterimi.

Diğer taraftan, belirtmek gerekir ki; bazı sıcaklıklarda yerel tuzak oluşumları gözlemlenebilmektedir. Ayrıca, protein zinciri üzerinde birbirine yakın olan amino asitler arasında hidrojen bağı oluşturmak, birbirine uzak olan amino asitlere göre entropik bakımdan çok daha kolaydır. Bu nedenlerden dolayı, modelimizin tamamıyla β -yapılı ve $\alpha + \beta$ yapılı proteinlerin katlanma dinamiğinde ve katlı yapı tahmininde de yüksek başarı göstermesini sağlayacak yeni enerji terimlerine geliştirilmesine veya mevcut terimlerin iyileştirilmesine ihtiyaç duyulabilecektir.

Geliştirdiğimiz modelin bir diğer önemli başarısı; IenH proteini için $T_{sim} = 3.1$ 'de elde edilen katlanma ve açılma geçişlerinin ara durum konformasyonlarında herhangi bir relaksasyon olmaksızın "ya hep ya hiç" şeklinde oldukça kooperatif şekilde gerçekleşmesidir (bkz. Şekil 4.31). Serbest enerji profili incelendiğinde; katlı yapıdan açık yapıya (tersi de doğru) geçiş yapabilmek için $2k_B T$ yüksekliğinde bir bariyerin aşılması gerektiği gözlemlenmektedir.

Diğer taraftan, konformasyonel değişimler sırasında hasıl olan yapısal geçişler ve dolayısıyla enerjideki dalgalanmalara bağlı ölçülen ısı kapasitesinin sıcaklığa bağlı değişimi incelendiğinde; ısı kapasitesinin yüksek bir maksimum veya tepe noktası değerine sahip olduğu görülecektir. Ayrıca, deneysel termodinamik kooperativite parametresi κ hesaplandığında değeri 0.88 olarak bulunmuştur. κ 'nın bu değeri, IenH



Şekil 4.31 : Ienh proteininin $T_{sim} = 3.1$ 'de elde edilen E ve RMSD trajektörileri, bu trajektörden elde edilen serbest enerji profili ve ısı kapasitesi değişimi.

proteinin deneylerde gözlemlendiği gibi iki durum termodinamik kooperativite davranışı gösterdiğini söylemektedir.

Mevcut literatür bilgimiz ışığında, şunu belirtebiliriz ki; bu tez kapsamında tüm-atom atom temsiline sahip, transfer edilebilir enerji terimleri ihtiva eden bir modelden ilk defa deneysel gözlemlerle uyumlu iki-durum termodinamik kooperativite davranışı gözlemlenmiştir.

5. TARTIŞMA ve SONUÇ

Bu tez çalışması kapsamında, moleküler dinamik simülasyon yöntemi ile protein içi etkileşim ağlarının katlanma termodinamiği ve kinetiği üzerinde olan etkisi detaylı bir şekilde incelenmiştir. İncelediğimiz 12 protein deneysel olarak iki-durum termodinamik ve kinetik davranışları göstermektedir. Aynı proteinler ele alınarak, protein içi etkileşim ağlarının protein termodinamiği üzerinde ki etkisi incelendiğinde; özellikle α -yapılı proteinler düşük r_{ese} veya OES değerlerinde kooperatif olmayan termodinamik davranışlar sergilemektedirler. Bununla birlikte, protein ailesi farklılığı gözetmeksizin, r_{ese} ve dolayısıyla OES değerinin artması ile birlikte kooperatif olmayan katlanma/açılma davranışı sergileyen proteinlerin, yeterli etkileşim yoğunluğuna eriştiklerinde kooperatif iki-durum davranış sergiledikleri görülmüştür. Sadece etkileşim ağlarındaki etkileşim sayısının artırılmasıyla kooperatif olmayan termodinamik davranıştan, kooperatif olan bir termodinamik davranışa geçiş dikkate değerdir. Çünkü, deneysel sonuçlarla uyumlu davranış gösterecek protein modelleri geliştirilmesinde; protein içi etkileşim ağlarının doğru tanımlanmasının ne kadar önemli olduğunu ispat etmektedir. Diğer bir deyişle, proteinlerdeki etkileşim ağlarının tanımı, elde edilecek hal değişiminin tipini veya derecesini belirleyebilmektedir. İki-durum termodinamik davranış birinci derece (süreksiz, sigmoidal veya "ya hep ya hiç") hal değişimini betimlerken, çok düşük serbest enerji bariyerli veya bariyersiz (downhill) hal değişimi ise kooperatif olmayan ikinci derece (sürekli) hal değişimine karşılık gelmektedir.

Deneysel olarak bariyersiz davranış gösteren 1bbl proteini için dahi, simülasyon çalışmalarında fiziko kimyasal olarak kabul edilebilir sınırların üzerinde bir $r_{ese} = 8.66\text{\AA}$ mesafe kriteri kullanıldığı durumda, iki-durum davranış gösterebildiği görülmüştür. 1bbl proteininin diğer 12 proteinden çok daha yüksek r_{ese} (farklı bölgede) değerlerinde iki-durum davranışına geçiş göstermesi, kullandığımız modelin deneysel sonuçlar ile örtüşen sonuçlar verdiğini göstermektedir.

Özellikle kısa r_{ese} mesafelerinde amino asitlerin OES değerlerinde yüksek standart sapmaların olduğu gözlemlenmiştir. r_{ese} mesafesinin artması bu standart sapmanın

azalmasını sağlamaktadır. Böylelikle r_{ese} değeri arttıkça, proteinin sahip olduğu amino asitlerin daha homojen bağ veya etkileşme dağılımlarına sahip olduğu gösterilmiştir.

Etkileşim ağlarının protein katlanma kinetiğine etkisini incelediğimizde; aynı protein için farklı etkileşim ağlarında katlanma oranları arasında $10^{4.17} \approx 14,800$ kat gibi muazzam farklılık olduğu görülmüştür. Bu gözlem, etkileşim ağlarının serbest enerji profilleri üzerinde ki etkisi ile karşılaştırıldığında artan serbest enerji bariyer yüksekliği ile tutarlıdır. Bir proteinde sadece etkileşim ağlarını değiştirerek katlanma oranlarında bu derece yüksek farklılık gözlemlenmesi, etkileşim ağlarının protein kinetiği üzerinde baskın bir etkisi olduğunun kanıtı niteliğindedir.

Literatürde tanımlanan EDP parametreleri, aynı protein için tanımlanan farklı etkileşim ağları için elde ettiğimiz; katlanma oranlarındaki muazzam farklılığı yakalamakta veya betimlemekte son derece yetersiz kaldığı gözlemlenmiştir. Diğer bir deyişle, literatürde tanımlanan etkileşim düzen parametreleri (MEDP ve GEDP) protein ailelerinin sahip olduğu topolojik farklılıkları yakalayabildiği, fakat aynı protein topolojisi (yani aynı protein) içinde tanımlanan etkileşim ağlarının beraberinde getirdiği değişimleri veya farklılıkları yakalayamamaktadır.

Bu tez çalışmasında önerilen AEEDP etkileşim düzen parametresi ise; hem protein ailelerinin topolojik farklılıklarını hem de aynı protein topolojisi içindeki etkileşim ağlarının gösterdiği termodinamik hal değişimlerindeki farklılığı ve kinetik katlanma oranlarındaki muazzam farklılığı yakalamakta ve betimleyebilmektedir. Ayrıca diğer EDP parametrelerinde gözlemlenen farklı protein ailelerinin ayrışmasını gidererek, farklı protein aileleri için elde edilen termodinamik ve kinetik verilerin üst üste düşmesini sağlamıştır.

Geliştirdiğimiz AEEDP etkileşim düzen parametresi protein setimizdeki 12 protein ve her protein için tanımlanan 15 etkileşim ağından (yani, toplam $12 \times 15 = 180$) elde edilen simülasyondan katlanma oranları ile yüksek korelasyon katsayına $r = 0.97$ sahip olması; protein kinetikleri hakkında tahmin yapabilecek uygun bir araç olmasını sağlamaktadır. AEEDP parametresi simülasyon katlanma oranlarının yanı sıra deneysel katlanma oranları ile de yüksek korelasyon vermektedir.

AEEDP etkileşim düzen parametresinin termodinamik kooperativite ile olan ilişkisine bakıldığında; farklı ailelere mensup proteinleri bir bant aralığında toplanmasını sağlayarak,

bütün proteinlerin aynı AEEDP $\simeq 3.25$ eşik değerinde kooperatif davranış göstermeye başladığını görmekteyiz. Dolayısıyla, bariyersiz ve iki-durum davranışı gösteren proteinler için, protein içi etkileşim ağının ne derece doğru oluşturulup oluşturulmadığı AEEDP parametresi hesaplanarak test edilebilecektir.

Sonuç olarak, geliştirdiğimiz AEEDP etkileşim düzen parametresi diğer EDP parametrelerinin aksine; hem basitleştirilmiş zincir temsilli moleküler dinamik Gō modelinden elde edilen simülasyon sonuçları ile hem de 12 protein için var olan deneysel sonuçlar ile uyumludur. Daha da önemlisi, tek AEEDP etkileşim düzen parametresi ile proteinlerin hem termodinamik hem de kinetik davranışlarının eş zamanlı olarak betimlenebildiği bulunmuştur.

Tez çalışmasında ayrıca tüm atomların temsil edildiği, Monte Carlo protein modeli için transfer edilebilir istatistiksel etkileşim potansiyelleri geliştirilmiştir. Bu istatistiksel potansiyeller PISCES ile tespit edilen %50 zincir özdeşliği, $2.5\text{\AA}$ ve daha küçük kristal yapı çözünürlüğü, zincir uzunluğu 40 amino asit'den fazla toplamda 25,801 protein yapısı incelenerek yapılmıştır.

Atom çiftleri için arasındaki etkileşimleri, atom çiftlerinin ait oldukları amino asitlerin birbirlerine protein zinciri üzerindeki mesafesine ($d = 1, 2, 3, 4, 5, 6$ ve $d > 6$) göre ayrı ayrı sınıflandırılarak etkileşim potansiyelleri oluşturulduğunda; aynı atom çiftleri için elde edilen minimum enerji değerleri ve sayıları, minimum enerji değerlerine karşılık gelen mesafe değerlerinde inanılmaz farklılaşma gözlemlenmiştir. Sanki, aynı atom çiftinin zincir üzerindeki mesafesi farklı olduğunda, tamamıyla farklı atom çifti gibi davranmaktadır. Bu farklılıkların sebebi; dengede ki sistemlerin konformasyon tercihleri, atom veya amino asitlerin farklı fiziko-kimyasal özellikleri, birbirlerine kovalent bağ ile bağlı atomların sahip oldukları geometrik kısıtlılık veya atom çiftlerinin zincir üzeri yakınlık derecesinden kaynaklanmaktadır.

Geliştirdiğimiz $N-H \cdots O$ transfer edilebilir istatistiksel hidrojen bağı enerji teriminde, bir mesafe ve dört açı kriteri kullanılarak, hidrojen bağının yönelimsel ve düzlemsel özellikleri sağlanmıştır. Mevcut simülasyon sonuçlarımız hidrojen bağı enerji teriminin daha da spesifikleştirilebileceğini işaret etmektedir.

Protein zinciri üzerinde ardışık (yerel) olan amino asit üçlüleri için 7 tane farklı transfer edilebilir istatistiksel etkileşim potansiyeli tasarlanmıştır. Bu enerji terimleri

ile katlı yapı enerjisinin en düşük enerjili yapı olmasına, konformasyon uzayının düzgün örneklenmesine, tuzak konformasyon durumlarının engellenmesini ve yerel konformasyonların doğru bir şekilde betimlenmesi sağlanmıştır. Protein setimizde tamamıyla α -yapılı ve tamamıyla β -yapılı protein aileleri ayrıştırılarak en sık gözlemlenen ilk 100 amino asit üçlüleri ele alınarak aralarındaki farklılıklar incelenmiştir. Ele alınan bu amino asit üçlülerinde, amino asitlerin gözlemlenme sıklığının protein ailelerine göre değişimine bakıldığında; amino asitlerin proteinlerde gelişi güzel bir şekilde değil, tam aksine her protein ailesinde belirli amino asit üçlülerinin içinde çok daha yüksek sıklıkla olacak şekilde bulunmayı tercih ettiği gözlemlenmiştir. Bu nedenle, geliştirdiğimiz amino asit üçlüsü enerji terimlerinin protein dinamiğinde de belirleyici rol oynadığı aşikardır.

Amino asit üçlüsü enerji terimlerinin minimum değerlerine karşılık gelen parametre bin kombinasyonları incelendiğinde; minimum enerji uzayının çok kısıtlı olduğu gözlemlenmiştir. Bu durum protein katlanması modelleme çalışmalarında istenilen ve aranan bir durumdur. Diğer taraftan, örnek olarak sunulan spesifik amino asit üçlüleri için gözlemlenen tüm parametre bin kombinasyonları (yani, tüm parametre uzayı) karşılaştırıldığında, her amino asit üçlüsünde aynı amino asit üçlüsü enerji terimi için gözlemlenen parametre bin kombinasyonlarının farklılaştığı ve gözlemlenme popülasyonunun da farklılaştığı gözlemlenmiştir. Bu gözlem, amino asitlerin dahil oldukları amino asit üçlülerine göre ele alınmasının son derece önemli olduğunun bir başka kanıtıdır.

Geliştirilen 11 tane farklı transfer edilebilir istatistiksel etkileşim potansiyeli, Monte Carlo simülasyon kodumuz kullanılarak; pdb kodları 1bdd ve 1enh olan tamamıyla α -yapılı iki farklı protein için test edilmiştir. Bu iki protein için simülasyondan elde edilen katlanmış yapıları, deneylerden elde edilen yapılarla karşılaştırdığımızda; 1bdd proteini için 2.95Å, 1enh proteini için ise 2.29Å RMSD değerleri bulunmuştur. Aynı kıyaslamayı gevşetilmiş-katlı yapı ile tekrarladığımızda, RMSD değerlerini 1bdd proteini için 2.03Å ve 1enh proteini için 1.27Å olarak bulmaktayız. Bu derece küçük RMSD değerlerine sahip katlı yapıların elde edilmesi; geliştirilen enerji terimlerinin protein topolojisini ayırt etmekte, ikincil yapıların ve ikincil yapılar arasındaki etkileşmelerin (yani, üçüncül yapıların) doğru oluşmasını sağlamakta çok başarılı olduğunu göstermektedir.

Ayrıca elde edilen simülasyon sonuçlarında; açılma ve katlanma geçişlerinin kooperatif bir şekilde, herhangi bir relaksasyon olmadan (çok düşük ara-durum popülasyonu ile) gerçekleştiği görülmüştür. Ienh proteini için elde edilen serbest enerji profili analiz edildiğinde; katlı yapıdan açık yapıya (veya tersi de doğru) geçiş yapabilmek için $2k_B T$ yüksekliğinde bir bariyerin aşılması gerektiği bulunmuştur. Aynı protein için deneysel termodinamik kooperativite parametresi κ hesaplandığında, değeri 0.88 olarak bulunmuştur. Bu değer, Ienh proteininin deneylerde gözlemlendiği gibi geliştirdiğimiz protein modelinde de iki durum kooperativite davranışı gösterdiğini işaret etmektedir. Mevcut literatür taramamıza göre, altı çizilerek belirtilmesi gerekir ki; tüm-atom temsiline sahip ve transfer edilebilir etkileşim potansiyelleri içeren bir protein modelinde; deneysel gözlemler ile uyumlu iki durum termodinamik kooperativite davranışı ilk defa geliştirdiğimiz bu model yardımıyla elde edilmiştir.

KAYNAKLAR

- [1] **Moore, P.B., Hendrickson, W.A., Henderson, R. ve Brunger, A.T.** (2022). The protein-folding problem: Not yet solved, *Science (New York, N.Y.)*, 375(6580), 507.
- [2] **Dill, K.A., Ozkan, S.B., Shell, M.S. ve Weikl, T.R.** (2008). The Protein Folding Problem, *Annual Review of Biophysics*, 37(1), 289–316, <http://www.annualreviews.org/doi/10.1146/annurev.biophys.37.092707.153558>, 721 citations (Crossref) [2022-08-16].
- [3] **Sosnick, T.R. ve Barrick, D.** (2011). The folding of single domain proteins—have we reached a consensus?, *Current Opinion in Structural Biology*, 21(1), 12–24.
- [4] **Galzitskaya, O.** (2010). Is protein folding rate dependent on number of folding stages? Modeling of protein folding with ferredoxin-like fold, *Biochemistry (Moscow)*, 75(6), 717–727, <http://journals.scholarsportal.info/detailsundefined,00000>.
- [5] **Knott, M. ve Chan, H.S.** (2006). Criteria for downhill protein folding: Calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity, *PROTEINS: Structure, Function, and Bioinformatics*, 65(2), 373–391, <http://onlinelibrary.wiley.com/doi/10.1002/prot.21066/full>, 00000.
- [6] **Matthews, C.R.**, (1987). [26] Effect of point mutations of the folding of globular proteins, *Methods in Enzymology*, cilt154 of *Recombinant DNA Part E*, Academic Press, s.498–511, <https://www.sciencedirect.com/science/article/pii/0076687987540927>.
- [7] **Chan, H.S., Kaya, H. ve Shimizu, S.** (2002). Computational methods for protein folding: scaling a hierarchy of complexities, *Current topics in computational molecular biology*, 403–447, <http://books.google.com/books?hl=en&lr=&id=MraXBw48KKYC&oi=fnd&pg=PA403&dq=%22for+protein+structure+prediction+require+various+degrees+of+understanding%22+%22predict+protein+behaviors+from+elementary+physical+forces.+At+present,%22+%22become+increasingly+important+for+providing+structural+and+functional%22+&ots=NaWCm3AMV5&sig=toHwrSUmET0WRlkyBuyETtRtEic>.
- [8] **Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. ve Karplus, M.** (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *Journal of*

Computational Chemistry, 4(2), 187–217, <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540040211>, [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540040211](https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540040211).

- [9] **Xu, D. ve Zhang, Y.** (2012). Ab Initio Protein Structure Assembly Using Continuous Structure Fragments and Optimized Knowledge-based Force Field, *Proteins*, 80(7), 1715–1735, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3370074/>.
- [10] **Duan, Y. ve Kollman, P.A.** (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science (New York, N.Y.)*, 282(5389), 740–744.
- [11] **Lindorff-Larsen, K., Maragakis, P., Piana, S. ve Shaw, D.E.** (2016). Picosecond to Millisecond Structural Dynamics in Human Ubiquitin, *The Journal of Physical Chemistry B*, 120(33), 8313–8320, <https://pubs.acs.org/doi/10.1021/acs.jpcb.6b02024>, 82 citations (Crossref) [2022-08-16].
- [12] **Sarkar, A., Gasic, A.G., Cheung, M.S. ve Morrison, G.** (2022). Effects of Protein Crowders and Charge on the Folding of Superoxide Dismutase 1 Variants: A Computational Study, *The Journal of Physical Chemistry B*, 126(24), 4458–4471, <https://doi.org/10.1021/acs.jpcb.2c00819>, publisher: American Chemical Society.
- [13] **Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E. ve Kolinski, A.** (2016). Coarse-Grained Protein Models and Their Applications, *Chemical Reviews*, 116(14), 7898–7936, <https://doi.org/10.1021/acs.chemrev.6b00163>, publisher: American Chemical Society.
- [14] **Piana, S., Klepeis, J.L. ve Shaw, D.E.** (2014). Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations, *Current Opinion in Structural Biology*, 24, 98–105, <https://linkinghub.elsevier.com/retrieve/pii/S0959440X13002157>, 364 citations (Crossref) [2022-08-16].
- [15] **Kendrew, J.C.** (1961). The three-dimensional structure of a protein molecule, *Scientific American*, 205, 96–110.
- [16] **Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. ve Phillips, D.C.** (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis, *Nature*, 181(4610), 662–666, <https://www.nature.com/articles/181662a0>, number: 4610 Publisher: Nature Publishing Group.
- [17] **Levinthal, C.** (1968). Are there pathways for protein folding?, *Journal de Chimie Physique*, 65, 44–45, <https://jcp.edpsciences.org/articles/jcp/abs/1968/01/jcp196865p44/jcp196865p44.html>, publisher: EDP Sciences.
- [18] **Anfinsen, C.B.** (1973). Principles that govern the folding of protein chains, *Science (New York, N.Y.)*, 181(4096), 223–230.

- [19] **Pennisi, E.** (2012). ENCODE Project Writes Eulogy for Junk DNA, *Science*, 337(6099), 1159–1161, <https://www.science.org/doi/10.1126/science.337.6099.1159>.
- [20] **Bragg, L.** (1976). The development of X-ray analysis, *Contemporary Physics*, 17(1), 103–104, <https://doi.org/10.1080/00107517608210844>, publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/00107517608210844>.
- [21] **Baldwin, E.T., Weber, I.T., St Charles, R., Xuan, J.C., Appella, E., Yamada, M., Matsushima, K., Edwards, B.F., Clore, G.M. ve Gronenborn, A.M.** (1991). Crystal structure of interleukin 8: symbiosis of NMR and crystallography., *Proceedings of the National Academy of Sciences*, 88(2), 502–506, <https://www.pnas.org/doi/abs/10.1073/pnas.88.2.502>, publisher: Proceedings of the National Academy of Sciences.
- [22] **Cheng, Y.** (2015). Single-Particle Cryo-EM at Crystallographic Resolution, *Cell*, 161(3), 450–457.
- [23] **Bank, P.D.** (1971). Protein data bank, *Nature New Biol*, 233, 223.
- [24] **Dana, J.M., Gutmanas, A., Tyagi, N., Qi, G., O'Donovan, C., Martin, M. ve Velankar, S.** (2019). SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins, *Nucleic Acids Research*, 47(D1), D482–D489, <https://doi.org/10.1093/nar/gky1114>.
- [25] **Liu, H., Zhao, Y. ve Sun, J.** (2023). Heterogeneous Nucleation in Protein Crystallization, *Biomimetics*, 8(1), 68, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9944892/>.
- [26] **Taketomi, H., Ueda, Y. ve Gō, N.** (1975). Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions, *International Journal of Peptide and Protein Research*, 7(6), 445–459, 00000.
- [27] **Ueda, Y., Taketomi, H. ve Gō, N.** (1978). Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme, *Biopolymers*, 17(6), 1531–1548, <http://onlinelibrary.wiley.com/doi/10.1002/bip.1978.360170612/abstract>, 00267.
- [28] **Go, N. ve Taketomi, H.** (1979). Studies on protein folding, unfolding and fluctuations by computer simulation. III. Effect of short-range interactions, *International Journal of Peptide and Protein Research*, 13(3), 235–252, 00026.
- [29] **Go, N. ve Taketomi, H.** (1979). Studies on protein folding, unfolding and fluctuations by computer simulation. IV. Hydrophobic interactions, *International Journal of Peptide and Protein Research*, 13(5), 447–461, 00037 bibtex.
- [30] **Jackson, S.E.** (1998). How do small single-domain proteins fold?, *Folding & Design*, 3(4), R81–91.

- [31] **Jackson, S.E. ve Fersht, A.R.** (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry*, 30(43), 10428–10435.
- [32] **Plaxco, K.W., Simons, K.T. ve Baker, D.** (1998). Contact order, transition state placement and the refolding rates of single domain proteins, *Journal of molecular biology*, 277(4), 985–994, <http://www.sciencedirect.com/science/article/pii/S002228369891645X>, 00000.
- [33] **Kaya, H. ve Chan, H.S.** (2003). Origins of Chevron Rollovers in Non-Two-State Protein Folding Kinetics, *Physical Review Letters*, 90(25), 258104, <http://link.aps.org/doi/10.1103/PhysRevLett.90.258104>, 00047.
- [34] **Kaya, H., Liu, Z. ve Chan, H.S.** (2005). Chevron Behavior and Isostable Enthalpic Barriers in Protein Folding: Successes and Limitations of Simple Gō-like Modeling, *Biophysical Journal*, 89(1), 520–535, <http://www.sciencedirect.com/science/article/pii/S0006349505727004>.
- [35] **Sieradzan, A.K., Krupa, P., Scheraga, H.A., Liwo, A. ve Czaplewski, C.** (2015). Physics-based potentials for the coupling between backbone- and side-chain-local conformational states in the UNited RESidue (UNRES) force field for protein simulations, *Journal of Chemical Theory and Computation*, 11(2), 817–831.
- [36] **Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., Labonte, J.W., Pacella, M.S., Bonneau, R., Bradley, P., Dunbrack, R.L., Das, R., Baker, D., Kuhlman, B., Kortemme, T. ve Gray, J.J.** (2017). The Rosetta all-atom energy function for macromolecular modeling and design, *Journal of chemical theory and computation*, 13(6), 3031–3048, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5717763/>, 566 citations (Crossref) [2022-08-16] 00270.
- [37] **Kaya, H. ve Chan, H.S.** (2002). Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding?, *Journal of Molecular Biology*, 315(4), 899–909, <http://linkinghub.elsevier.com/retrieve/pii/S0022283601952660>, 00077.
- [38] **Clementi, C., Nymeyer, H. ve Onuchic, J.N.** (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins, *Journal of Molecular Biology*, 298(5), 937–953.
- [39] **Hu, J., Chen, T., Wang, M., Chan, H.S. ve Zhang, Z.** (2017). A critical comparison of coarse-grained structure-based approaches and atomic models of protein folding, *Physical Chemistry Chemical Physics*, 19(21), 13629–13639, <http://xlink.rsc.org/?DOI=C7CP01532A>, 23 citations (Crossref) [2022-08-16].
- [40] **Kaya, H. ve Chan, H.S.** (2003). Solvation Effects and Driving Forces for Protein Thermodynamic and Kinetic Cooperativity: How Adequate is

Native-centric Topological Modeling?, *Journal of Molecular Biology*, 326(3), 911–931, <http://www.sciencedirect.com/science/article/pii/S0022283602014341>, 00162.

- [41] **Kaya, H., Uzunoğlu, Z. ve Chan, H.S.** Spatial ranges of driving forces are a key determinant of protein folding cooperativity and rate diversity, 88(4), <http://link.aps.org/doi/10.1103/PhysRevE.88.044701>.
- [42] **Privalov, P.L. ve Khechinashvili, N.N.** A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study, 86(3), 665–684, <http://www.sciencedirect.com/science/article/pii/S0022283674901880>, 00000.
- [43] **Sivaraman, T., Arrington, C.B. ve Robertson, A.D.** (2001). Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin, *Nature Structural Biology*, 8(4), 331–333.
- [44] **Meisner, W.K. ve Sosnick, T.R.** (2004). Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: Implications for downhill folding, *Proceedings of the National Academy of Sciences of the United States of America*, 101(44), 15639–15644.
- [45] **Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. ve Lindahl, E.** GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, 1–2, 19–25, <http://www.sciencedirect.com/science/article/pii/S2352711015000059>.
- [46] **Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L. ve Schulten, K.** Scalable molecular dynamics with NAMD, 26(16), 1781–1802, <http://onlinelibrary.wiley.com/doi/10.1002/jcc.20289/abstract>.
- [47] **Plimpton, S.** Fast parallel algorithms for short-range molecular dynamics, 117(1), 1–19, <http://www.sciencedirect.com/science/article/pii/S002199918571039X>.
- [48] **Glaser, J., Nguyen, T.D., Anderson, J.A., Lui, P., Spiga, E., Millan, J.A., Morse, D.C. ve Glotzer, S.C.** Strong scaling of general-purpose molecular dynamics simulations on GPUs, 192, 97–107, <http://www.sciencedirect.com/science/article/pii/S0010465515000867>.
- [49] **Ryckaert, J.P., Ciccotti, G. ve Berendsen, H.J.** (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics*, 23(3), 327–341, <https://linkinghub.elsevier.com/retrieve/pii/S0021999177900985>.
- [50] **Verlet, L.** (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Physical Review*, 159(1), 98–103, <https://link.aps.org/doi/10.1103/PhysRev.159.98>.

- [51] **Veitshans, T., Klimov, D. ve Thirumalai, D.** (1997). Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties, *Folding and Design*, 2(1), 1–22, <http://www.sciencedirect.com/science/article/pii/S1359027897000023>, 00367.
- [52] **Hockney, R.W., Goel, S.P. ve Eastwood, J.W.** (1974). Quiet high-resolution computer models of a plasma, *Journal of Computational Physics*, 14(2), 148–158, <https://www.sciencedirect.com/science/article/pii/0021999174900102>.
- [53] **Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. ve Bourne, P.E.** (2000). The Protein Data Bank, *Nucleic Acids Research*, 28(1), 235–242, <https://academic.oup.com/nar/article/28/1/235/2384399>, 24342 citations (Crossref) [2022-08-16] 29944.
- [54] **Hamelryck, T. ve Manderick, B.** (2003). PDB file parser and structure class implemented in Python, *Bioinformatics*, 19(17), 2308–2310, <https://academic.oup.com/bioinformatics/article/19/17/2308/205793/> PDB-file-parser-and-structure-class-implemented-in, 00105.
- [55] **Sato, S., Religa, T.L., Daggett, V. ve Fersht, A.R.** (2004). Testing protein-folding simulations by experiment: B domain of protein A, *Proceedings of the National Academy of Sciences*, 101(18), 6952–6956, <https://www.pnas.org/doi/full/10.1073/pnas.0401396101>, publisher: Proceedings of the National Academy of Sciences.
- [56] **Robien, M.A., Clore, G.M., Omichinski, J.G., Perham, R.N., Appella, E., Sakaguchi, K. ve Gronenborn, A.M.** (1992). Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*, *Biochemistry*, 31(13), 3463–3471, <https://doi.org/10.1021/bi00128a021>, publisher: American Chemical Society.
- [57] **Mueller, U., Perl, D., Schmid, F.X. ve Heinemann, U.** (2000). Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein11Edited by D. C. Rees, *Journal of Molecular Biology*, 297(4), 975–988, <https://www.sciencedirect.com/science/article/pii/S0022283600936027>.
- [58] **Luisi, D.L., Kuhlman, B., Sideras, K., Evans, P.A. ve Raleigh, D.P.** (1999). Effects of varying the local propensity to form secondary structure on the stability and folding kinetics of a rapid folding mixed / protein: characterization of a truncation mutant of the N-terminal domain of the ribosomal protein L911Edited by P. E. Wright, *Journal of Molecular Biology*, 289(1), 167–174, <https://www.sciencedirect.com/science/article/pii/S0022283699927420>.

- [59] **Hanaoka, S., Nagadoi, A., Yoshimura, S., Aimoto, S., Li, B., de Lange, T. ve Nishimura, Y.** (2001). NMR structure of the hrap1 myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of myb DNA-binding domains11Edited by P. E. Wright, *Journal of Molecular Biology*, 312(1), 167–175, <https://www.sciencedirect.com/science/article/pii/S0022283601949241>.
- [60] **Furukawa, K., Oda, M. ve Nakamura, H.** (1996). A small engineered protein lacks structural uniqueness by increasing the side-chain conformational entropy, *Proceedings of the National Academy of Sciences*, 93(24), 13583–13588, <https://www.pnas.org/doi/full/10.1073/pnas.93.24.13583>, publisher: Proceedings of the National Academy of Sciences.
- [61] **Fazi, B., Cope, M.J.T.V., Douangamath, A., Ferracuti, S., Schirwitz, K., Zucconi, A., Drubin, D.G., Wilmanns, M., Cesareni, G. ve Castagnoli, L.** (2002). Unusual Binding Properties of the SH3 Domain of the Yeast Actin-binding Protein Abp1: STRUCTURAL AND FUNCTIONAL ANALYSIS *, *Journal of Biological Chemistry*, 277(7), 5290–5298, [https://www.jbc.org/article/S0021-9258\(19\)82562-0/abstract](https://www.jbc.org/article/S0021-9258(19)82562-0/abstract), publisher: Elsevier.
- [62] **Jiménez, M.A., Villegas, V., Santoro, J., Serrano, L., Vendrell, J., Avilés, F.X. ve Rico, M.** (2003). NMR solution structure of the activation domain of human procarboxypeptidase A2, *Protein Science*, 12(2), 296–305, <https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.0227303>, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1110/ps.0227303>.
- [63] **Feng, S., Chen, J.K., Yu, H., Simon, J.A. ve Schreiber, S.L.** (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions, *Science (New York, N.Y.)*, 266(5188), 1241–1247.
- [64] **Noble, M.E., Musacchio, A., Saraste, M., Courtneidge, S.A. ve Wierenga, R.K.** (1993). Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin, *The EMBO journal*, 12(7), 2617–2624.
- [65] **Sato, S., Religa, T.L., Daggett, V. ve Fersht, A.R.** (2004). Testing protein-folding simulations by experiment: B domain of protein A, *Proceedings of the National Academy of Sciences*, 101(18), 6952–6956, <https://www.pnas.org/doi/full/10.1073/pnas.0401396101>, publisher: Proceedings of the National Academy of Sciences.
- [66] **Ferguson, N., Sharpe, T.D., Schartau, P.J., Sato, S., Allen, M.D., Johnson, C.M., Rutherford, T.J. ve Fersht, A.R.** (2005). Ultra-fast Barrier-limited Folding in the Peripheral Subunit-binding Domain Family, *Journal of Molecular Biology*, 353(2), 427–446, <https://www.sciencedirect.com/science/article/pii/S002228360500968X>.
- [67] **McPhalen, C.A. ve James, M.N.G.** (1987). Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds, *Biochemistry*, 26(1),

261–269, <https://doi.org/10.1021/bi00375a036>, publisher: American Chemical Society.

- [68] **Wikstroem, M., Drakenberg, T., Forsen, S., Sjoerding, U. ve Bjoerck, L.** (1994). Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G, *Biochemistry*, 33(47), 14011–14017, <https://doi.org/10.1021/bi00251a008>, publisher: American Chemical Society.
- [69] **Ivankov, D.N., Garbuzynskiy, S.O., Alm, E., Plaxco, K.W., Baker, D. ve Finkelstein, A.V.** (2003). Contact order revisited: influence of protein size on the folding rate, *Protein Science: A Publication of the Protein Society*, 12(9), 2057–2062.
- [70] **Gromiha, M. ve Selvaraj, S.** (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction11Edited by P. E. Wright, *Journal of Molecular Biology*, 310(1), 27–32, <http://linkinghub.elsevier.com/retrieve/pii/S0022283601947758>, 00000.
- [71] **PEARSON, K.** (1920). NOTES ON THE HISTORY OF CORRELATION, *Biometrika*, 13(1), 25–45, <https://doi.org/10.1093/biomet/13.1.25>.
- [72] **Bross, P., Corydon, T.J., Andresen, B.S., Jørgensen, M.M., Bolund, L. ve Gregersen, N.** (1999). Protein misfolding and degradation in genetic diseases, *Human Mutation*, 14(3), 186–198, <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291098-1004%281999%2914%3A3%3C186%3A%3AAID-HUMU2%3E3.0.CO%3B2-J>, 143 citations (Crossref) [2022-08-16].
- [73] **Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. ve Hassabis, D.** (2021). Highly accurate protein structure prediction with AlphaFold, *Nature*, 596(7873), 583–589, <https://www.nature.com/articles/s41586-021-03819-2>, number: 7873 Publisher: Nature Publishing Group.
- [74] **Zheng, W., Zhang, C., Li, Y., Pearce, R., Bell, E.W. ve Zhang, Y.** (2021). Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations, *Cell Reports Methods*, 1(3), 100014, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8336924/>.
- [75] **Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. ve Baker, D.** (2020). Improved protein structure prediction using predicted interresidue

orientations, *Proceedings of the National Academy of Sciences of the United States of America*, 117(3), 1496–1503, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6983395/>.

- [76] **Zheng, W., Li, Y., Zhang, C., Zhou, X., Pearce, R., Bell, E.W., Huang, X. ve Zhang, Y.** (2021). Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14, *Proteins*, 89(12), 1734–1751.
- [77] **MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D. ve Karplus, M.** (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins, *The Journal of Physical Chemistry. B*, 102(18), 3586–3616.
- [78] **Wang, G. ve Dunbrack, Jr, R.L.** (2003). PISCES: a protein sequence culling server, *Bioinformatics*, 19(12), 1589–1591, <https://doi.org/10.1093/bioinformatics/btg224>.
- [79] **Shortle, D.** (2003). Propensities, probabilities, and the Boltzmann hypothesis, *Protein Science : A Publication of the Protein Society*, 12(6), 1298–1302, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2323900/>.
- [80] **Sippl, M.J.** Calculation of Conformational Ensembles from Potentials of Mean Force, 25.
- [81] **Vaiana, S.M., Manno, M., Emanuele, A., Palma-Vittorelli, M.B. ve Palma, M.U.** The Role of Solvent in Protein Folding and in Aggregation.
- [82] **Ingold, C.K.** (1940). The Nature of the Chemical Bond and the Structure of Molecules and Crystals, *Nature*, 145(3678), 644–645, <https://www.nature.com/articles/145644b0>, number: 3678 Publisher: Nature Publishing Group.
- [83] **Dill, K.A.** (1990). Dominant forces in protein folding, *Biochemistry*, 29(31), 7133–7155, <https://doi.org/10.1021/bi00483a001>, publisher: American Chemical Society.
- [84] **Stickle, D.F., Presta, L.G., Dill, K.A. ve Rose, G.D.** Hydrogen Bonding in Globular Proteins.
- [85] **Cserző, M. ve Simon, I.** (1989). Regularities in the primary structure of proteins, *International Journal of Peptide and Protein Research*, 34(3), 184–195, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1399-3011.1989.tb00229.x>, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1399-3011.1989.tb00229.x>.
- [86] **Reeck, G.R. ve Fisher, L.** (1973). A Statistical Analysis of the Amino Acid Compositions of Proteins, *International Journal of Peptide and Protein Research*, 5(2), 109–117, <https://onlinelibrary.wiley.com/>

doi/abs/10.1111/j.1399-3011.1973.tb02325.x, _eprint:
<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1399-3011.1973.tb02325.x>.

- [87] **Fleming, P.R., Milne, M.D., Morgan, D.G. ve Morris, M.** (1972). A statistical analysis of amino acid sequence in protein chains, *International Journal of Bio-Medical Computing*, 3(2), 91–100, <https://www.sciencedirect.com/science/article/pii/0020710172900402>.
- [88] **Skipper, L.**, (2005). PROTEINS | Overview, **P. Worsfold, C. Poole, A. Townshend ve M. Miró**, (düzenleyenler), *Encyclopedia of Analytical Science* (Third Edition), Academic Press, Oxford, s.412–419, <https://www.sciencedirect.com/science/article/pii/B978008101983200493X>.
- [89] **Lu, H. ve Skolnick, J.** (2001). A distance-dependent atomic knowledge-based potential for improved protein structure selection, *Proteins: Structure, Function, and Genetics*, 44(3), 223–232, <https://onlinelibrary.wiley.com/doi/10.1002/prot.1087>.
- [90] **Ramachandran, G.N., Ramakrishnan, C. ve Sasisekharan, V.** (1963). Stereochemistry of polypeptide chain configurations, *Journal of Molecular Biology*, 7, 95–99.
- [91] **Binder, K. ve Heermann, D.W.** (2010). *Monte Carlo Simulation in Statistical Physics: An Introduction*, cilt 0 of *Graduate Texts in Physics*, Springer Berlin Heidelberg, Berlin, Heidelberg, <https://link.springer.com/10.1007/978-3-642-03163-2>.
- [92] **Kussell, E., Shimada, J. ve Shakhnovich, E.I.** (2002). A structure-based method for derivation of all-atom potentials for protein folding, *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), 5343–5348, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC122771/>.
- [93] **Hubner, I.A., Deeds, E.J. ve Shakhnovich, E.I.** (2005). High-resolution protein folding with a transferable potential, *Proceedings of the National Academy of Sciences*, 102(52), 18914–18919, <https://www.pnas.org/doi/10.1073/pnas.0502181102>, publisher: Proceedings of the National Academy of Sciences.
- [94] **Coutsias, E.A., Seok, C., Jacobson, M.P. ve Dill, K.A.** (2004). A kinematic view of loop closure, *Journal of Computational Chemistry*, 25(4), 510–528.
- [95] **Yang, J.S., Chen, W.W., Skolnick, J. ve Shakhnovich, E.I.** (2007). All-Atom Ab Initio Folding of a Diverse Set of Proteins, *Structure*, 15(1), 53–63, <https://linkinghub.elsevier.com/retrieve/pii/S0969212606004746>.
- [96] **Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. ve Teller, E.** (1953). Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, 21(6), 1087–1092,

[https://pubs.aip.org/jcp/article/21/6/1087/202680/
Equation-of-State-Calculations-by-Fast-Computing.](https://pubs.aip.org/jcp/article/21/6/1087/202680/Equation-of-State-Calculations-by-Fast-Computing)

EKLER

Moleküler Dinamik Simülasyon Sonuçları

Tablo A.1 : Basitleştirilmiş zincir modeli ile moleküler dinamik simülasyonu ile incelenen protein içi etkileşim ağlarının simülasyon parametreleri ve çıktıları.

pdb	L	r_{ese}	$N_{bağısız}^{kath}$	T_{sim}	$T_{geçiş} = T_{sim}(C_P^{max})$	κ	$\log_{10} k_f^{sim}$
1C9O	66	3.250	32	0.405	0.404	0.43	-5.488
1C9O	66	3.500	53	0.522	0.521	0.51	-5.689
1C9O	66	3.570	61	0.561	0.560	0.58	-5.708
1C9O	66	3.724	70	0.604	0.602	0.79	-5.966
1C9O	66	3.750	79	0.642	0.642	0.82	-6.007
1C9O	66	4.000	106	0.768	0.766	1.00	-6.767
1C9O	66	4.250	127	0.856	0.854	1.00	-7.253
1C9O	66	4.500	139	0.906	0.904	1.00	-7.363
1C9O	66	4.750	153	0.954	0.953	1.00	-7.365
1C9O	66	5.000	163	0.992	0.990	1.00	-7.259
1C9O	66	5.250	178	1.056	1.054	0.99	-7.280
1C9O	66	5.500	190	1.098	1.096	0.97	-7.062
1C9O	66	5.750	203	1.153	1.151	1.00	-7.838
1C9O	66	6.000	218	1.207	1.208	1.01	-8.229
1C9O	66	6.400	247	1.318	1.319	1.01	-8.344
1CQU	39	3.250	40	0.580	0.575	0.39	-5.052
1CQU	39	3.500	54	0.697	0.696	0.63	-5.407
1CQU	39	3.580	57	0.717	0.717	0.75	-5.495
1CQU	39	3.700	60	0.743	0.740	0.85	-5.535
1CQU	39	3.750	63	0.765	0.761	1.00	-5.731
1CQU	39	4.000	71	0.827	0.825	0.99	-5.742
1CQU	39	4.250	81	0.894	0.894	0.99	-5.980
1CQU	39	4.500	91	0.960	0.956	1.00	-6.230
1CQU	39	4.750	97	1.006	1.005	0.99	-6.309
1CQU	39	5.000	111	1.105	1.096	1.00	-6.530
1CQU	39	5.100	115	1.123	1.119	1.00	-6.699
1CQU	39	5.500	123	1.188	1.177	1.01	-6.970
1CQU	39	5.550	125	1.186	1.185	1.01	-6.886
1CQU	39	6.000	142	1.315	1.302	1.00	-7.170
1CQU	39	6.400	164	1.455	1.446	1.01	-7.300
1FEX	59	3.250	42	0.450	0.455	0.33	-4.803
1FEX	59	3.500	56	0.548	0.551	0.37	-5.023
1FEX	59	3.750	63	0.603	0.604	0.39	-5.039
1FEX	59	4.000	70	0.639	0.642	0.43	-5.117
1FEX	59	4.250	79	0.705	0.704	0.52	-5.318
1FEX	59	4.500	87	0.747	0.746	0.54	-5.343
1FEX	59	4.750	100	0.790	0.791	0.55	-5.403
1FEX	59	5.000	115	0.872	0.872	0.70	-5.631

Tablo A.1 : (devamı)

pdb	L	r_{ese}	$N_{\text{bağısız}}^{\text{kath}}$	T_{sim}	$T_{\text{geçiş}} = T_{\text{sim}}(C_P^{\text{max}})$	κ	$\log_{10} k_f^{\text{sim}}$
1FEX	59	5.360	135	0.962	0.960	0.74	-5.638
1FEX	59	5.450	139	0.984	0.982	0.90	-5.752
1FEX	59	5.500	143	1.005	1.004	0.95	-5.770
1FEX	59	5.530	145	1.013	1.012	0.98	-5.694
1FEX	59	6.000	165	1.105	1.102	1.00	-6.080
1FEX	59	6.400	185	1.218	1.216	0.99	-6.050
1IDY	54	3.250	32	0.491	0.435	0.36	-4.402
1IDY	54	3.500	41	0.551	0.506	0.31	-4.585
1IDY	54	3.750	47	0.565	0.569	0.36	-4.869
1IDY	54	4.000	60	0.680	0.679	0.43	-5.105
1IDY	54	4.250	66	0.717	0.718	0.49	-5.152
1IDY	54	4.500	69	0.737	0.738	0.50	-5.214
1IDY	54	4.750	78	0.785	0.784	0.51	-5.208
1IDY	54	4.897	86	0.823	0.822	0.55	-5.323
1IDY	54	5.000	93	0.850	0.850	0.51	-5.260
1IDY	54	5.160	100	0.876	0.874	0.54	-5.279
1IDY	54	5.500	107	0.935	0.929	0.58	-5.340
1IDY	54	5.674	113	0.967	0.961	0.64	-5.500
1IDY	54	5.793	117	0.991	0.987	0.63	-5.419
1IDY	54	5.840	122	1.031	1.027	0.72	-5.481
1IDY	54	5.889	126	1.060	1.057	0.89	-5.605
1IDY	54	6.000	128	1.072	1.068	1.00	-5.710
1IDY	54	6.190	140	1.138	1.135	1.00	-5.848
1IDY	54	6.400	147	1.182	1.172	1.00	-5.850
1JO8	58	3.250	37	0.435	0.435	0.39	-5.405
1JO8	58	3.500	60	0.570	0.570	0.54	-5.697
1JO8	58	3.550	66	0.599	0.597	0.60	-5.759
1JO8	58	3.600	71	0.627	0.626	0.64	-5.884
1JO8	58	3.700	80	0.679	0.677	0.78	-6.070
1JO8	58	3.750	85	0.708	0.704	0.95	-6.201
1JO8	58	4.000	109	0.832	0.831	1.00	-6.998
1JO8	58	4.250	127	0.913	0.913	1.00	-6.953
1JO8	58	4.500	137	0.961	0.961	1.01	-7.185
1JO8	58	4.750	149	1.016	1.015	1.01	-7.359
1JO8	58	5.000	164	1.085	1.085	1.01	-8.185
1JO8	58	5.250	177	1.141	1.140	1.01	-7.531
1JO8	58	5.500	184	1.170	1.171	1.01	-7.570
1JO8	58	6.000	209	1.277	1.279	1.01	-8.000
1JO8	58	6.400	235	1.390	1.393	1.01	-8.520
1O6X	81	3.250	50	0.465	0.463	0.40	-5.471
1O6X	81	3.500	72	0.580	0.579	0.59	-5.728
1O6X	81	3.570	76	0.590	0.589	0.60	-5.810
1O6X	81	3.650	83	0.622	0.622	0.65	-5.874
1O6X	81	3.750	90	0.657	0.656	0.87	-5.950
1O6X	81	4.000	103	0.713	0.713	0.99	-6.097

Tablo A.1 : (devamı)

pdb	L	r_{ese}	$N_{\text{bağısız}}^{\text{kath}}$	T_{sim}	$T_{\text{geçiş}} = T_{\text{sim}}(C_P^{\text{max}})$	κ	$\log_{10} k_f^{\text{sim}}$
1O6X	81	4.250	118	0.780	0.781	1.00	-6.428
1O6X	81	4.500	137	0.855	0.854	1.01	-6.838
1O6X	81	4.750	153	0.911	0.911	1.01	-7.178
1O6X	81	5.000	169	0.963	0.964	1.01	-7.672
1O6X	81	5.250	188	1.028	1.027	1.01	-7.874
1O6X	81	5.500	204	1.087	1.088	1.01	-8.300
1O6X	81	6.000	234	1.183	1.185	1.01	-8.521
1O6X	81	6.400	276	1.337	1.335	1.01	-8.581
1RLQ	56	3.250	31	0.450	0.449	0.43	-5.494
1RLQ	56	3.500	48	0.550	0.549	0.62	-5.909
1RLQ	56	3.650	55	0.597	0.594	0.72	-6.148
1RLQ	56	3.700	61	0.623	0.620	0.75	-5.980
1RLQ	56	3.720	64	0.635	0.634	0.89	-6.115
1RLQ	56	3.750	66	0.645	0.643	0.91	-6.193
1RLQ	56	4.000	81	0.730	0.728	1.00	-6.563
1RLQ	56	4.250	94	0.801	0.801	1.01	-6.746
1RLQ	56	4.500	108	0.873	0.872	1.00	-6.737
1RLQ	56	4.750	123	0.940	0.938	1.00	-7.020
1RLQ	56	5.000	132	0.981	0.980	1.00	-7.144
1RLQ	56	5.250	145	1.039	1.037	1.01	-7.436
1RLQ	56	5.500	160	1.098	1.103	1.01	-7.642
1RLQ	56	6.000	184	1.210	1.212	1.01	-8.090
1RLQ	56	6.400	210	1.330	1.329	1.01	-8.402
1SHF	59	3.250	34	0.440	0.440	0.35	-5.294
1SHF	59	3.500	53	0.520	0.523	0.35	-5.355
1SHF	59	3.750	75	0.661	0.662	0.47	-5.639
1SHF	59	3.900	82	0.710	0.709	0.63	-5.798
1SHF	59	4.000	93	0.773	0.771	0.89	-5.980
1SHF	59	4.250	115	0.876	0.874	1.00	-6.910
1SHF	59	4.500	129	0.944	0.943	1.01	-7.107
1SHF	59	4.750	139	0.984	0.985	1.01	-7.348
1SHF	59	5.000	146	1.018	1.016	1.01	-7.477
1SHF	59	5.250	158	1.072	1.070	1.01	-7.313
1SHF	59	5.500	170	1.123	1.125	1.01	-7.509
1SHF	59	6.000	203	1.260	1.260	1.01	-7.801
1SHF	59	6.400	227	1.367	1.364	1.01	-8.046
1SS1	60	3.250	48	0.530	0.536	0.34	-4.946
1SS1	60	3.500	63	0.640	0.639	0.42	-5.091
1SS1	60	3.750	69	0.680	0.681	0.45	-5.216
1SS1	60	4.000	74	0.724	0.722	0.51	-5.298
1SS1	60	4.250	80	0.748	0.749	0.54	-5.361
1SS1	60	4.500	90	0.786	0.784	0.56	-5.350
1SS1	60	4.750	106	0.848	0.847	0.69	-5.517
1SS1	60	4.820	111	0.877	0.876	0.71	-5.501
1SS1	60	4.870	116	0.900	0.898	0.81	-5.582

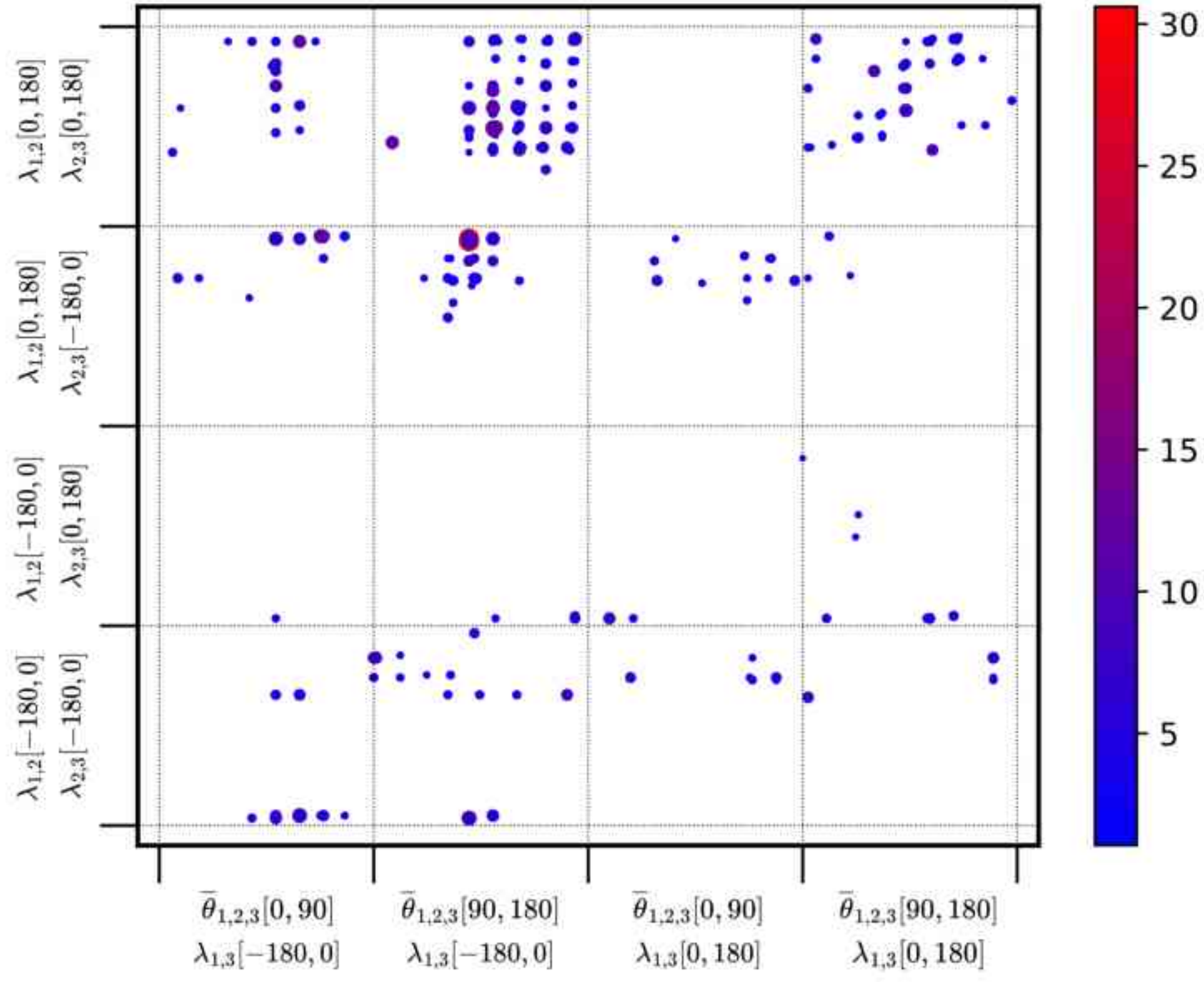
Tablo A.1 : (devamı)

pdb	L	r_{ese}	$N_{\text{bağısız}}^{\text{kath}}$	T_{sim}	$T_{\text{geçiş}} = T_{\text{sim}}(C_P^{\text{max}})$	κ	$\log_{10} k_f^{\text{sim}}$
1SS1	60	5.000	120	0.915	0.914	1.01	-5.717
1SS1	60	5.250	134	0.982	0.982	1.00	-5.711
1SS1	60	5.500	152	1.076	1.078	1.00	-5.818
1SS1	60	5.750	169	1.173	1.173	0.99	-5.924
1SS1	60	6.000	183	1.250	1.249	1.00	-6.240
1SS1	60	6.400	206	1.363	1.358	1.00	-6.520
1W4J	51	3.250	36	0.511	0.512	0.38	-4.888
1W4J	51	3.500	41	0.557	0.553	0.42	-4.920
1W4J	51	3.750	50	0.645	0.645	0.52	-5.257
1W4J	51	4.000	56	0.700	0.698	0.58	-5.397
1W4J	51	4.250	64	0.755	0.753	0.58	-5.335
1W4J	51	4.500	69	0.790	0.789	0.60	-5.370
1W4J	51	4.750	84	0.880	0.879	0.52	-5.258
1W4J	51	5.000	96	0.933	0.931	0.57	-5.386
1W4J	51	5.250	103	0.967	0.967	0.60	-5.332
1W4J	51	5.271	107	0.987	0.986	0.68	-5.470
1W4J	51	5.350	112	1.019	1.017	0.76	-5.571
1W4J	51	5.380	113	1.020	1.017	0.76	-5.548
1W4J	51	5.453	116	1.035	1.030	0.90	-5.591
1W4J	51	5.500	120	1.067	1.061	1.00	-5.800
1W4J	51	5.670	124	1.097	1.095	1.01	-6.018
1W4J	51	5.970	137	1.167	1.169	1.01	-6.340
1W4J	51	6.000	138	1.181	1.178	1.01	-6.380
1W4J	51	6.400	165	1.368	1.368	1.01	-6.970
2CI2	64	3.250	34	0.380	0.394	0.34	-5.565
2CI2	64	3.500	51	0.510	0.504	0.48	-5.516
2CI2	64	3.750	72	0.620	0.621	0.71	-5.696
2CI2	64	3.850	79	0.653	0.652	0.94	-5.852
2CI2	64	3.930	86	0.680	0.681	0.98	-6.065
2CI2	64	4.000	97	0.733	0.733	1.00	-6.464
2CI2	64	4.250	112	0.803	0.802	1.01	-6.937
2CI2	64	4.500	131	0.880	0.879	1.01	-6.981
2CI2	64	4.750	138	0.915	0.914	1.01	-7.439
2CI2	64	5.000	153	0.966	0.963	1.01	-7.772
2CI2	64	5.250	163	1.006	1.005	1.01	-7.888
2CI2	64	5.500	173	1.042	1.042	1.01	-7.558
2CI2	64	5.750	191	1.119	1.119	1.01	-7.563
2CI2	64	6.000	208	1.190	1.187	1.01	-8.300
2CI2	64	6.400	240	1.327	1.329	1.01	-9.105
2PTL	62	3.250	46	0.490	0.502	0.43	-5.222
2PTL	62	3.500	64	0.610	0.609	0.73	-5.787
2PTL	62	3.600	70	0.629	0.628	0.80	-5.884
2PTL	62	3.650	78	0.670	0.668	1.00	-6.100
2PTL	62	3.750	87	0.715	0.713	1.01	-6.463
2PTL	62	4.000	103	0.779	0.776	1.00	-6.636

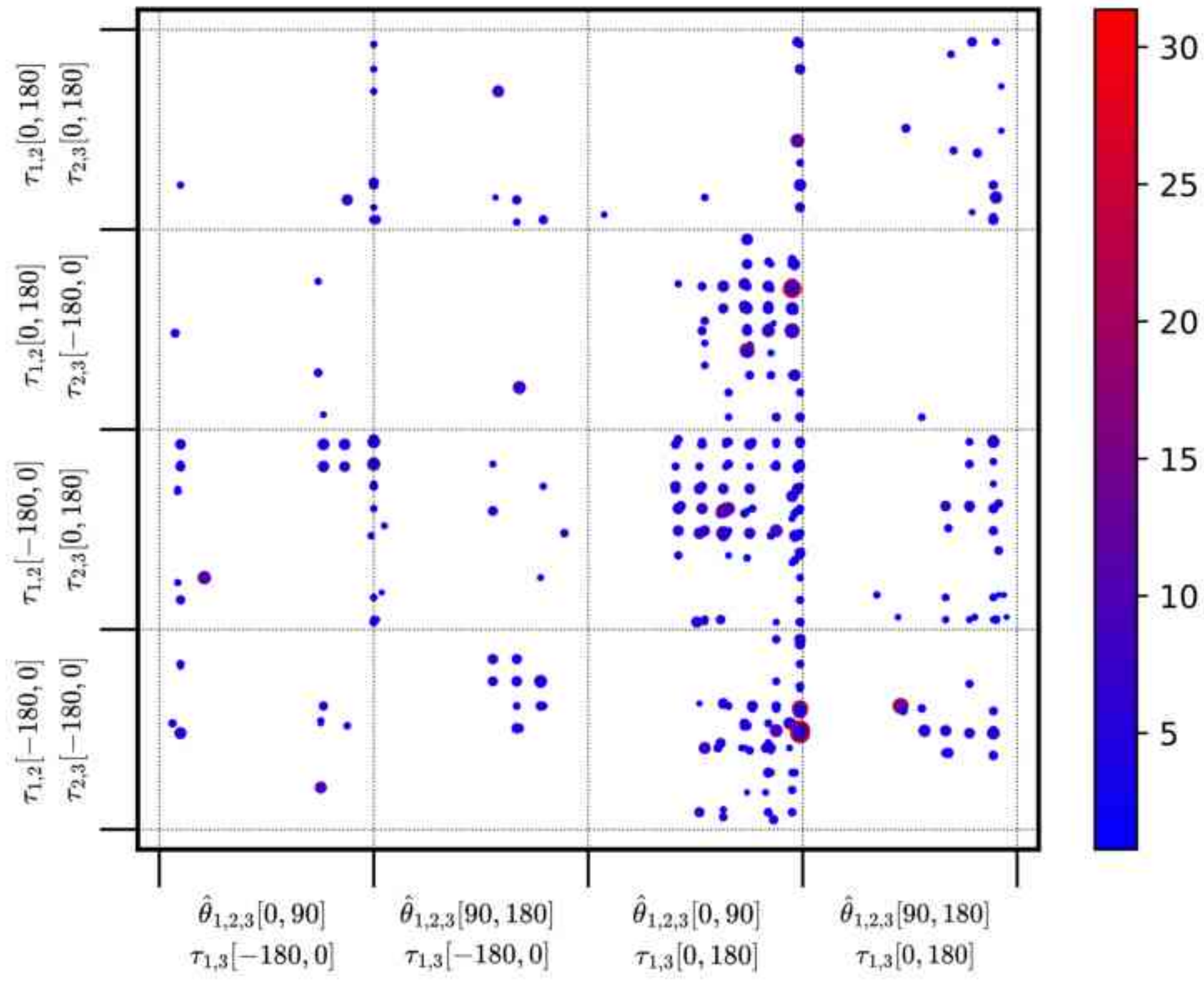
Tablo A.1 : (devamı)

pdb	L	r_{ese}	$N_{\text{bağısız}}^{\text{kath}}$	T_{sim}	$T_{\text{geçiş}} = T_{\text{sim}}(C_P^{\text{max}})$	κ	$\log_{10} k_f^{\text{sim}}$
2PTL	62	4.250	111	0.812	0.809	1.01	-6.690
2PTL	62	4.500	120	0.853	0.852	1.01	-6.872
2PTL	62	4.750	130	0.896	0.895	1.01	-6.971
2PTL	62	5.000	138	0.930	0.925	1.01	-7.008
2PTL	62	5.250	157	1.013	1.012	1.01	-7.539
2PTL	62	5.500	175	1.095	1.094	1.01	-7.773
2PTL	62	5.750	197	1.185	1.187	1.01	-8.459
2PTL	62	6.000	216	1.274	1.271	1.01	-8.983
2PTL	62	6.400	249	1.407	1.408	1.01	-9.389
1BBL	37	3.250	10	0.440	0.449	0.35	-3.669
1BBL	37	3.500	13	0.460	0.484	0.33	-3.937
1BBL	37	3.750	17	0.495	0.535	0.33	-4.014
1BBL	37	4.000	24	0.528	0.551	0.36	-4.081
1BBL	37	4.250	31	0.580	0.628	0.34	-4.299
1BBL	37	4.500	39	0.625	0.650	0.33	-4.586
1BBL	37	4.750	45	0.630	0.641	0.36	-4.665
1BBL	37	5.000	57	0.732	0.741	0.36	-4.710
1BBL	37	5.250	61	0.767	0.768	0.38	-4.706
1BBL	37	5.500	65	0.789	0.792	0.37	-4.702
1BBL	37	6.000	74	0.849	0.851	0.37	-4.735
1BBL	37	6.400	83	0.952	0.956	0.38	-4.841
1BBL	37	6.718	93	1.039	1.046	0.42	-4.922
1BBL	37	6.997	101	1.103	1.118	0.39	-4.963
1BBL	37	7.269	109	1.182	1.201	0.46	-5.056
1BBL	37	7.388	117	1.286	1.285	0.58	-5.160
1BBL	37	7.605	125	1.350	1.351	0.80	-5.398
1BBL	37	7.887	133	1.417	1.415	0.72	-5.283
1BBL	37	8.108	141	1.511	1.498	0.60	-5.266
1BBL	37	8.212	149	1.561	1.560	0.72	-5.301
1BBL	37	8.350	157	1.624	1.621	0.92	-5.506
1BBL	37	8.447	165	1.696	1.693	0.94	-5.648
1BBL	37	8.582	173	1.754	1.750	0.95	-5.642
1BBL	37	8.657	181	1.815	1.813	0.98	-5.877
1BBL	37	8.800	189	1.868	1.870	0.99	-6.092
1BBL	37	9.064	197	1.948	1.953	0.99	-6.017
1BBL	37	9.226	205	1.993	1.996	1.00	-6.160
1BBL	37	9.309	213	2.061	2.061	1.00	-6.319
1BBL	37	9.498	221	2.122	2.127	1.00	-6.212
1BBL	37	9.664	229	2.190	2.193	1.00	-6.445

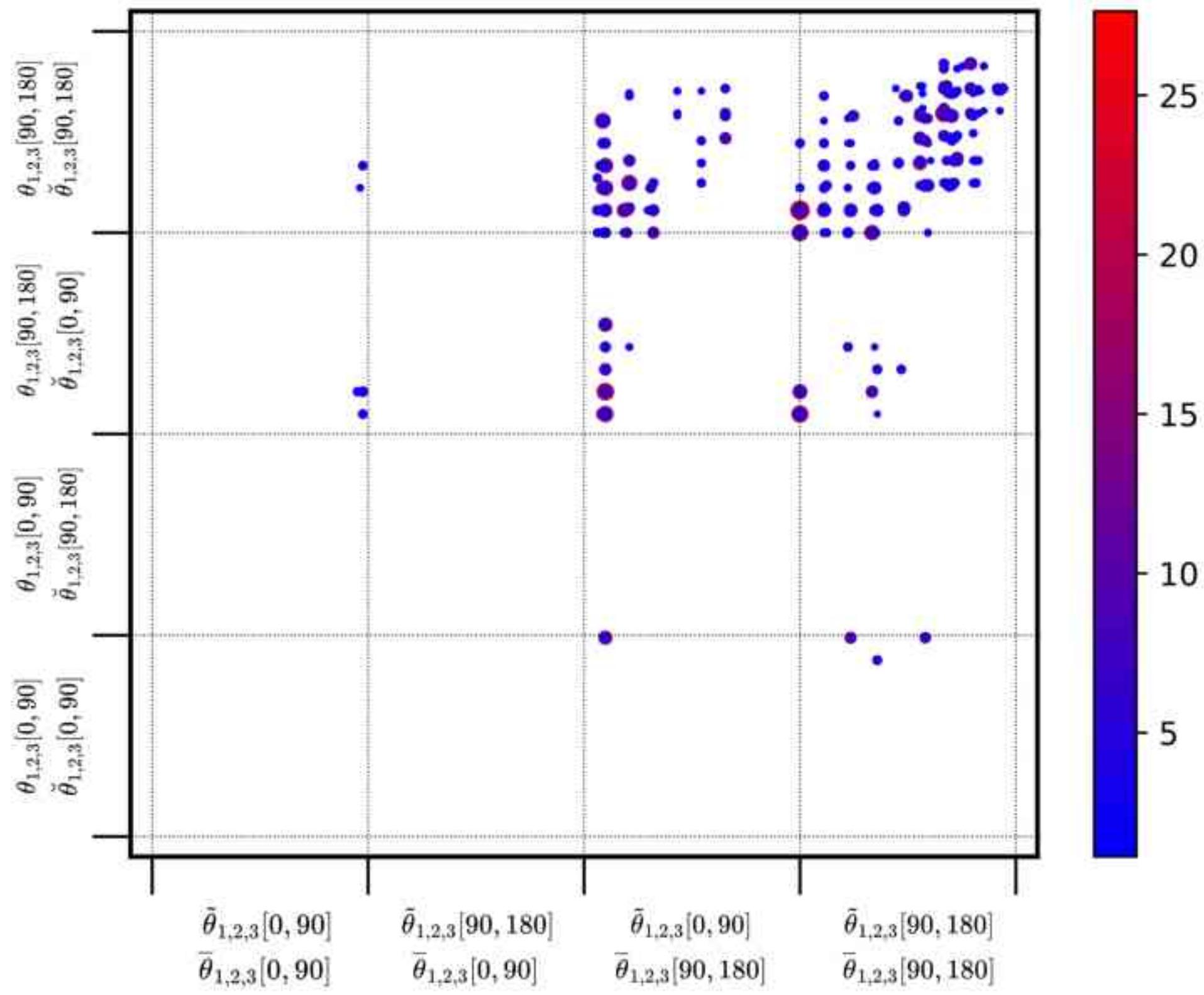
Zincir Üzerinde Yerel Olan Amino Asit Üçlüleri için Minimum Enerji Grafikleri



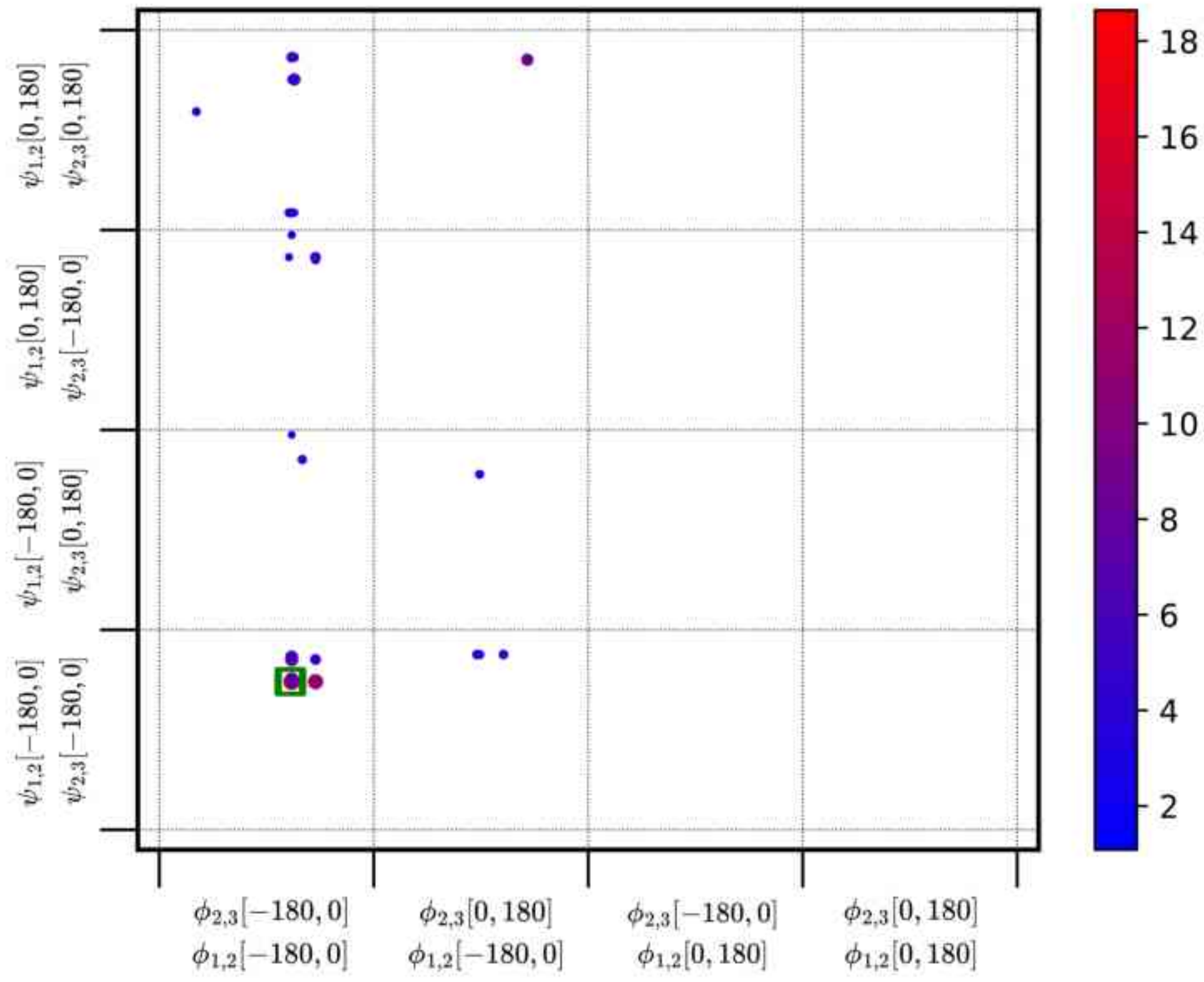
Şekil B.1 : Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\lambda-\bar{\theta}}((a_1, a_2, a_3)_i, \lambda_{21}, \lambda_{22}, \lambda_{23}, \bar{\theta}_{123})$ enerji fonksiyonunun minimum değere sahip olduğu λ_{21} , λ_{22} , λ_{23} ve $\bar{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi



Şekil B.2 : Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\tau-\hat{\theta}} \left((a_1, a_2, a_3)_i, \tau_{21}, \tau_{22}, \tau_{23}, \hat{\theta}_{123} \right)$ enerji fonksiyonunun minimum değere sahip olduğu τ_{21} , τ_{22} , τ_{23} ve $\hat{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi



Şekil B.3 : Tüm amino asit üçlüleri için $E_{\text{üçlü}}^{\theta} \left((a_1, a_2, a_3)_i, \theta_{123}, \check{\theta}_{123}, \tilde{\theta}_{123}, \bar{\theta}_{123}, \hat{\theta}_{123} \right)$ enerji fonksiyonunun minimum değere sahip olduğu θ_{123} , $\check{\theta}_{123}$, $\bar{\theta}_{123}$ ve $\hat{\theta}_{123}$ parametrelerinin bin kombinasyonlarının gösterimi



Şekil B.4 : Ortanca amino asitin ALA olduğu durumdaki amino asit üçlüleri için $E_{\text{üçlü}}^{\psi-\phi}((a_1, a_2, a_3), \psi_{12}, \phi_{12}, \psi_{23}, \phi_{23})$ enerji fonksiyonunun minimum değere sahip olduğu ψ_{12} , ϕ_{12} , ψ_{23} ve ϕ_{23} parametrelerinin bin kombinasyonlarının gösterimi

ÖZGEÇMİŞ

yılında doğdu. Darüşşafaka lisesinin bitirdi. Gaziantep Üniversitesi, Fizik Mühendisliği bölümünden mezun oldu. Gaziantep Üniversitesi, Tıbbi Biyoloji Anabilim Dalı, Biyobilişim ve Hesaplamalı Biyoloji yüksek lisans programından mezun oldu.