

Title	データマイニングによる塩水濃度と水溶液の動的構造の関連性に関する研究
Author(s)	中村, 大城
Citation	
Issue Date	2012-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/10470
Rights	
Description	Supervisor:DAM HIEU CHI 准教授, 知識科学研究科, 修士

修 士 論 文

データマイニングによる塩水濃度と
水溶液の動的構造の関連性に関する研究

北陸先端科学技術大学院大学

知識科学研究科知識科学専攻

中村 大城

2012 年 3 月

修 士 論 文

データマイニングによる塩水濃度と 水溶液の動的構造の関連性に関する研究

指導教官 DAM HIEU CHI 准教授

審査委員主査 Ho Tu Bao 教授

審査委員 林幸雄 准教授

審査委員 由井蘭隆也 准教授

北陸先端科学技術大学院大学

知識科学研究科知識科学専攻

1050042 中村 大城

提出年月: 2012 年 2 月

目次

第1章 序論	3
1.1 はじめに	3
1.2 研究目的	4
第2章 研究背景	5
2.1 水	5
2.2 NaCl 水溶液	5
2.2.1 イオンと水分子	5
2.2.2 飽和濃度を超える NaCl 水溶液	8
2.3 計算機上における塩水の再現	8
2.4 水溶液の計算	8
第3章 実験手法	10
3.1 分子シミュレーション	10
3.1.1 分子動力学法	10
3.2 データマイニング	11
3.2.1 データマイニングのプロセス	12
3.2.2 前処理	13
3.2.3 特徴空間の設計	14
3.3 解析	17
第4章 データ作成	18
4.1 シミュレーション	18
4.1.1 モデル作成	18
4.1.2 計算	27

4.2 解析	27
4.2.1 水構造の解析	27
第 5 章 結果と考察	29
5.1 NaCl 濃度の影響	29
5.2 考察	40
5.2.1 液中濃度	40
5.2.2 飽和濃度以上の水溶液	43
第 6 章 結論	44
6.1 今後の課題	44
6.2 まとめ	44
謝辞	46
付 録 A Amber	47
A.1 leap	47
A.2 sander	47
参考文献	48

第1章 序論

1.1 はじめに

塩水は自然界に大量に存在する私たちのもっとも身近な水溶液の一つである。

身近にある塩水として代表的なのは海水である。海水は地球上における水分の97%を占めており、もっとも大量の塩水といえるだろう。私たち生命は海から生まれたといわれており、体液も塩水の要素を持っており医療や生体化学の面では重要な溶液となっている。

一方、工業的な面からも重要な溶液の一つである。溶液は金属の腐食に影響を与える [1]。そのため、前述の海水について腐食防止のための研究は盛んにおこなわれている。例えば、海水環境下で部材が滑り接触する機器においては、摺動部に腐食摩耗が発生し、機器の性能や寿命に悪影響を及ぼすことが知られている [2]。腐食摩耗と電気化学的腐食および両者の相互作用により発生し、機械的摩耗や科学的摩耗の場合に比べ、材料を大きく損傷させる場合がある。そのため、塩水に関して調査が進められている。

また、日常生活においても塩水は多くの場所で見ることができる。特に食の場においてはなくてはならない存在である。塩は基本的な味覚で、五味の一つである塩味を発現する味物質とされている [3,4]。昨今、より良い生活を追求する”Quality of Life”が話題になっているが、生きていく上で欠かせない食の満足を向上させるために塩味は重要な要素となっている。そのため、味覚に関する研究は各所で進められており [5,6]、擬似的に味覚を感知し、定量的に評価をおこなう味覚センサも開発されている [7,8]。

これらのように塩水についての研究を進めることは多くの意義を持つ。塩水を研究することで工業や医療、そして日常生活への貢献につながる。

現在、塩水の研究手法は様々な方法が用いられているが、その一つとして分子シミュレーションが挙げられる。分子シミュレーションは方程式を解くことでコンピュータ上に物理特性その特性を示し、解明する手法である。以前は計算速度の関係から設定できる系が小さく、短期間のものに限られていた。しかし、計算機の急速な発達により大規模かつ長時間のシミュレーションを実行することができるようになった。そして研究はさらに進んでいき、得られるデータは巨大なものになっていった。ところが、得られたデータの

解析方法は主に統計的手法が用いられ、以前からある手法のままである。

しかし、現状のシミュレーション解析方法は巨大データに対する有効なアプローチ方法をもっておらず、統計値や直感による解析しかおこなわれていない。そこで、巨大データの有効的に使用するため帰納法的アプローチにより新しい知識を得る手法の開発が望まれる。

巨大データの帰納法的アプローチとしてデータマイニングが挙げられる。この手法は現在、各所で巨大データの解析に用いられている。たとえば WEB では巨大な情報が複雑なネットワークを形成しており、世界中で研究の対象とされている。その成果として検索エンジンが挙げられる。これは大きな情報を解析する手法データマイニングを用いて設計されている。これは計算機に多量データから学習させることにより、内在する法則性を導き出す。

この、データマイニングを用いることでシミュレーションにより得られる巨大データを解析することで、従来と異なるアプローチで塩水の特性が明らかになる可能性がある。

1.2 研究目的

本研究ではデータマイニングを用いることで、塩水の分子をシミュレーションするにより得られる巨大データを解析する。それにより、シミュレーション結果に対し、帰納法的アプローチによる知識発見を目的とする。

第2章 研究背景

2.1 水

重要な物質「水」。私たちにとって身近な物質であるが、その表面はどのような構造になっているのか未だよくわかっておらず、未解決問題とされている。[9]。これまでの研究では、水の表面に存在している水分子同士は強い水素結合で結ばれていることが明らかにされたが、具体的な構造については分からず、研究者によるさまざまな仮説が存在している。よく知られている仮説の一つに、この強い水素結合を氷の構造と結び付けて「界面の水は氷とよく似た構造を持つ」というモデルが提案されました。しかし、理論計算や熱力学的な考察と相容れず、未だに世界的な論争が続いていました。

地球上のいたるところに存在し、生命に欠かすことのできない水は、私たちの身近な物質である。その分子は酸素原子 1 個と水素原子 2 個という単純な構成となっており、。しかし、その集合体は互いに酸素原子の孤立電子対と水素原子との間に形成される 4 つの水素結合により、特異な性質を示す。そのため、水分子の挙動は複雑な物となり、その性質を探ることは難問の一つとして知られている。

2.2 NaCl 水溶液

NaCl 水溶液は電解質の溶液であり、全ての Na および Cl が陽イオンの Na イオンと陰イオンの Cl イオンとして完全電離する。また、水分子と同程度のサイズである K イオン - 水分子間のクーロン力は水分子 - 水分子間に働く分子間力よりも 4 倍もの大きい値をとることから、イオンの周囲に存在する水分子の状態はイオンと水分子間に働くクーロン力の影響を強く受ける。

2.2.1 イオンと水分子

イオン周りの水分子の状態を決める要素の一つとして、何個の水分子がイオンに接しているかを示す配位数がある。この配位数は最大で 12 個までとることができ、イオンの種類や測定方法によって異なる

値を示すが、無機の単原子イオンの場合は 4 または 6 となる [10, 11]。本実験で用いた Na イオンの配位数は 4 である。

ある瞬間、水分子の半径よりも小さい陽イオンが四面体構造をとっている水分子の中心に入ったときを考える。水分子同士の正四面体配置は 10^{-12} 秒程度しか維持されないため、水分子同士の結合が切れると陽イオンと水分子間に働くクーロン力は水素結合より大きくなり、水分子が回転してイオンへ負の極を向ける。このときイオンから働く力が大きいので水分子も熱運動をしているが、純水中の水分子の熱運動に比べて回転運動および並進運動共に小さいものとなる。そのため、4 個の水分子が陽イオンの周囲にある間は水分子が負の極を陽イオンに向けた構造をとり、これを正の水和をとるという。

そして陽イオンのサイズを大きくしていくと、クーロン力は分子間距離の二乗に反比例することから力が次第に弱くなる。そしてイオン - 水分子間と水分子 - 水分子間に働く力が等しくなる。この場合、イオンは水分子に対して水分子と同じ作用をするのでこれらの水分子の挙動は純水中と同じようなものとなる。

そしてさらに陽イオン半径を大きくすると、イオン - 水分子間よりも水分子 - 水分子間に働く力のほうが大きくなる。すると、水分子が水素結合が切れて自由に回転できるようになった瞬間は陽イオンへ負の極を向けるが、水素結合をつくりやすい位置に飛んできた水分子と水素結合をつくる。その結果、陽イオンの周りの水分子の回転運動は純水中よりも激しいものとなり、水分子は陽イオンから離れやすくなる。これを負の水和という。

本実験で用いた Na イオンの大きさは水分子半径に比べて小さく、正の水和をとる。

2.2.1-a イオンの挙動

一方、イオンの挙動も水分子から影響を受ける。ストークスの法則によりイオンの速度はイオン半径と水の粘度に反比例する [12]。そのためこの速度を測るとことにより水中のイオン半径がわかり、 r_s で表す。一方、イオンの大きさを表す量として結晶半径 r_c がある。これはイオン結晶中のイオンを球体の剛体として求めた半径である。

それぞれのイオン半径を表 2.1 に示す。これより Li および Na イオンは水中ではイオン半径が大きくなり、K、Rb、Cl イオンは小さくなる。これはイオン周りの水分子の熱運動の状態が原因である。Li および Na イオンは正の水和をとっており、イオン周りに存在する水分子の動きは阻害される。したがって、イオンの挙動は数個の分子を伴って動くようにみえるためストークス半径が結晶半径より大きくなる。また、K、Rb、Cl イオンは負の水和をとり、イオン周辺の水分子は純水よりも動きやすい状態をとる。そのため、

表 2.1: イオンのストークスの半径 r_s および結晶半径 r_c [11]

	Li イオン	Na イオン	K イオン	Rb イオン	Cl イオン	Br イオン
$r_s(\text{\AA})$	2.38	1.84	1.25	1.18	1.25	1.18
$r_c(\text{\AA})$	0.60	0.95	1.33	1.48	1.81	1.95

イオンが動くときに発生する抵抗は小さくなり、ストークス半径は結晶半径より小さくなる。

これより、Na イオンは Cl イオンよりも移動速度が大きい。

2.2.2 飽和濃度を超える NaCl 水溶液

溶ける / 溶けないは、日常生活での感覚であり、それに応じて、溶液・溶質・溶媒という概念は、中学校で教えられる。溶液中の少数成分である溶質は、主要成分である溶媒に囲まれる。これが溶媒和であり、溶質 - 溶媒相互作用を通して、溶媒中での溶質の安定性、つまり、溶けやすさが決定される。そして溶質 - 溶媒界面では溶解と凝縮が同時に起こっており、安定的な状態では溶解速度と凝縮速度がつり合い平衡状態となっている。

NaCl の飽和濃度を超える場合、溶解した NaCl は析出する。そのため、水溶液の挙動だけでなく、固体-液体界面についての議論も必要になる。先行研究において、Na イオンは結晶に近づいていき、Cl イオンは結晶から離れていくことが報告されている [13]。これにより、イオンの挙動は結晶周辺において特徴的な挙動を示す。

2.3 計算機上における塩水の再現

NaCl は完全電離するイオンであり、水溶液中では単原子イオンとして存在する。水溶媒の中にこれを配置することでシミュレーションのモデルを作成する。

2.4 水溶液の計算

本来、物質の物性を調べるには実験装置を使い実験値を求める必要があるが、そのためには大掛かりな装置が必要になってしまう。

しかし近年ではパソコンの処理能力向上に伴い、計算機シミュレーションを利用することで一般的な実験手法では困難な条件下においても分子レベルで観察できる。

よって水溶液の構造を知るため、計算機シミュレーションを用いる。

計算機シミュレーションは、適当な分子間相互作用を仮定して、計算機上で溶液系を生成するものである。分子一つ一つを直接扱うことができるので、好みに応じてどのような情報を取り出すことも原理的に可能である [14–16]。

物質の構造や物性を評価・予測するために、分子シミュレーションを利用することは、近年特に盛んになっている [17]。それに伴い、シミュレーションの手法についても、その対象や目的から様々な方法が開発されており、なかでも分子動力学 (MD) 法、モンテカルロ (MC) 法、分子力学 (MM) 法などは広く用いら

れている。分子力学法が主に一分子の最も安定な構造を探索することを目的としているのに対し、モンテカルロ法や分子動力学法は多数の原子・分子から成る集合を、シミュレーションの主な対象としており、いくつかの解析法が研究されている [18, 19]。その中で、分子動力学法 (Molecular Dynamics) は物質を構成する原子や分子を古典力学の運動方程式に従って運動する質点あるいは剛体と見なして、その運動を時々刻々と迫っていくため、時間に依存した性質や振る舞いを調べることが可能である点が、大きな特長である。

第3章 実験手法

3.1 分子シミュレーション

分子シミュレーションとは、計算機を用いて物理現象や物質が持つ物性などを分子の動きを数値計算することにより解析する試みである。これを用いることにより分子一つ一つの変化が追跡可能になる。計算方法としては分子力学法やモンテカルト法などいくつかあるが、本研究ではその中の一つである分子動力学法を用いてシミュレーションを実行した。

3.1.1 分子動力学法

分子動力学法とは物質系を構成する一つ一つの原子や分子に対して運動方程式をコンピュータを用いて数値的に解き、位置や速度、エネルギーなどを時間変化に伴って追跡する方法である。

力を及ぼし合っている N 個の粒子の運動方程式は

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{F}_i(t) \quad (i = 1, 2, 3, \dots, N) \quad (3.1)$$

と表すことが出来る。ここで $\mathbf{r}_i$ と m_i は粒子 i の位置ベクトルおよび質量を示す。また、 $\mathbf{F}_i(t)$ は粒子 i が他の粒子から受ける力を示す。

この式 (3.1.1) を時間について積分することにより、任意の時間における各粒子の速度や位置を予測することが出来る。しかし、 $\mathbf{F}_i(t)$ が $\mathbf{r}_i$ の複雑な関数である場合、解を求めることが困難であるため差分近似法を用いる。

$\mathbf{r}_i$ を t でテイラー展開すると次式のようになる。

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \left(\frac{d\mathbf{r}_i(t)}{dt} \right) \Delta t + \frac{1}{2} \left(\frac{d^2 \mathbf{r}_i(t)}{dt^2} \right) \Delta t^2 + \dots \quad (3.2)$$

式 (3.2) より Δt は十分に短い時間のため、 Δt^3 以降を無視し、 $d\mathbf{r}_i(t)/dt$ を $\mathbf{v}_i(t)$ と表し、式 (3.1.1) から

$d^2\mathbf{r}_i(t)/dt^2$ として代入すると次式が得られる。

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{\mathbf{F}_i}{2m_i\Delta t^2} \quad (3.3)$$

これを解くことにより、時間変化に対して各原子の座標を得ることができる。

3.2 データマイニング

データマイニングはその名の通り、多量のデータから有用な知識を発掘する技術の総称である。まだ歴史は浅いが、すでに実績のある人工知能、機械学習、統計、情報検索などの技術を総合出来に活用した実用嗜好の情報処理技術であり、急速な進展を遂げている。

大容量記憶媒体の低価格化、計算機処理能力の向上、情報通信技術の急速な進展の効果があって、ネットワーク社会におけるデータ収集や活用が格段に容易になった。実際、計算機で処理できる情報の多くがインターネットを通して世界中に飛び交い、これらの情報に誰もがいつでも自由にアクセスできる時代に突入した。現在の WEB ページ数は 100 億ページにも達しており、多くの人が使っている検索エンジン Google は 8000 大の PC を使い、クラスタコンピューティングにより 7000 万件 / 日の検索を、1 件あたり 0.5 秒で処理している。このような情報化社会では情報の価値が極めて重要になり、量が質に変化する。データマイニングは時代の要請に応えるべく生まれてきた新しい技術である。

データマイニングはシーズとしての技術面からも、ニーズとしての応用面からも多くの注目を浴びている。20 年前のエキスパートシステムの到来、その後の過度の期待と失敗を思い出させるが、データマイニングはそれとは事情が異なり、情報化社会の基盤技術として定着するように思われる。両者とも知識を前面に出している点は共通であるが、前者は人間の専門家の頭にある知識の抽出と利用を目的にしているのに対し、後者はデータに内在する非明示的な知識を発掘し、利用しようとする点が違う。雑多な鉱石を含む鉱脈から貴重な金を掘り当てるのが難しいように、最初から掘り当てたいものがどこにあるのか、あるいは本当にあるのかわかっているわけではない。

データマイニングは valid(たまたま見つかったのではなく、汎化能力高く)、novel(まだ我々が知らない新規な物で)、useful(有効であって実際に使う事ができ)、understandable(我々に理解出来る) 知識を機械的に発掘するということが一般的に受け入れられている認識であるが、実際には至難の業である。この基準を満足するかどうかは、実際に知識を使う人でなければわからない。計算機で実際に出来る事は、数理的

なモデルを設定し、知識の表現形式と、データあるいは表現形式から計算できる知識の評価基準を決めて、評価値が高く、指定した形式に従った知識を探索することである。具体的には、データに内在する規則や特徴的パターンを発掘する。発掘のための技術としては、従来からアル統計やデータベースの技術のほかに、人工知能、特に機械学習の技術が融合して使われる。

データマイニングはネットワーク社会の到来がもたらせた情報洪水の諸問題に対する有望な解決策となることが期待されている。モデリングの観点から見れば、特定の目的のためだけに集められたものではない大量のデータから、目的に沿ったモデルを掘り起こす技術と見る事が出来る。

3.2.1 データマイニングのプロセス

知識発見 (KDD: Knowledge Discovery in Databases または Knowledge Discovery and Data Mining) あるいはデータマイニングと呼ばれる分野は、1990 年代初期に登場し、情報技術の進歩とデータ蓄積量の爆発的増大を背景に急速に発展してきた。データマイニングは膨大な量のデータに埋もれる興味深い知識の発見を目的としてデータベース管理、統計学、機械学習などを技術的基盤にもつ学際領域で会う [20–24]。

データマイニングは原則として、次に示す 5 つのステップで構成されるプロセスとして考えてよい。

1. 対象領域の理解とマイニング課題設定
2. データ収集とデータ前処理
3. データマイニング パターン / モデルの抽出
4. 発見した知識の解釈と評価 (結果の後処理)
5. 知識の実利用

狭義のデータマイニングは第 3 ステップに相当し、統計学や機械学習の理論に基づく各種マイニングのアルゴリズムを用いてデータからパターンやモデルを発見することをいう。高品質の発見を実現するには、第 1 ステップの課題設定、第 2 ステップのデータの収集・前処理の役割が大きい。

第 1 ステップは、マイニング対象データの所属領域 / 分野の理解とその文脈におけるマイニング課題の定式化をタスクとする。第 2 ステップでは、元になるデータの選定、データが不足する場合のデータ収集、そして集まったデータに対する適切な前処理がタスクである。前処理の目的は、データの基本的質を確保し、適切な量・質・形式のデータを次ステップへ提供することにある。データや課題に応じ、不良データ

除去、欠損値対策、属性選択やサンプリングによるデータ量の削減、データ変形に代表される処理を適宜適用するが、課題に対する洞察、データの特性とアルゴリズム等に関する知識が不可欠である。実際のデータマイニングでは、元データが多様な形式や質で提供されるため、マイニング・プロセス全体の7割以上の負荷を費やすことが多い。第3ステップで得られる特徴記述や予測パターン/モデルがそのままの状態では通用することはめったになく、第4ステップの後処理で適切な表現形式を与えられ、対象領域の文脈において専門家による解釈・評価を経ることで、実世界で活用される知識へとつながる。最終的に新規性が高く有用な知識を発見するまでには、利用者の満足のいく結果が得られるまで個別のステップあるいは一連のプロセスを繰り返す必要があり、この繰り返しが知識発見プロセスの本質ともいえる。第5ステップの発見した知識の実利用は知識発見の究極の目的でもある。

3.2.2 前処理

データマイニングのプロセスのうち、多くの研究がパターン発見アルゴリズムに集中している。しかし、マイニングの全プロセスにとっては、データの獲得、選択、前処理、変換がもっとも重要なものである。データあってのマイニングであり、マイニングのデータが悪ければ、良い知識は発掘できない。そのため、事前に次のような処理をおこなう。

- データマイニングアルゴリズムが処理できる形にデータを整形
- 型を一致させたり、欠損値を補完したりする

本研究ではここまでのプロセスを解析することで塩水の特性を調べる。

3.2.2-a データの整形

Amberによって得られる分子の挙動に関するデータは各ステップ毎に全てに原子の位置がどこに位置しているのかを示すものである。そのため、各原子がどのような挙動をとったのかを追うことはそのままのデータ形式では解析が困難である。そのため、データをステップ毎の位置ではなく、一つの原子がどのような軌道ををとりながら動くか示すようにデータを変換する。

3.2.2-b 欠陥値の補完

欠損値とは抜けているまたは異常なデータのことである。例えば人がデータ入力する場合は人的ミスなど。プログラムがデータを保存する場合は、プログラムのバグ、ハードディスクのエラーなどがあげられる。

欠損値の補完は主に次のような方法がある。

- 他のデータを見て、平均値や代表値などでデータで埋める。
- 該当するデータを削除する。
- 欠損値の多い属性を削除する。

どの補完方法を選ぶかはその時々に合わせてデータマイニングの実行者が経験などに基づいて判断しなくてはならない。もちろん、欠損値はなんでも補完すれば良いのではなく、意味のある欠損なのかを考え、また欠損してもよいデータ群なのかを検討する必要がある。

今回の計算では欠損値はなかったので特に処理はしていない。

3.2.3 特徴空間の設計

ある次元のデータ空間では、分類できないデータを次元を変えることで分類できるようになる場合がある。今回の場合では水分子挙動をシステム上解析するために示すように特徴空間をデザインする必要がある。

水分子の振る舞いは酸素原子 O の位置座標より抽出する。図 3.1、3.2、3.3 は典型的な水分子の軌道である。この軌道を分子の振る舞いとして扱い、定量化や特徴空間の構築をおこなう。

ある時間における軌道の平均位置（重心）を計算し、同じ時間での重心の移動量を特徴空間の要素として用いる。ある時間 $T = N\delta t$ の時間において、

$$\text{group1} : (x(t_i), y(t_i), z(t_i)), i = 1, 2, 3, \dots, N \quad \text{group2} : (x(t_i), y(t_i), z(t_i)), i = N + 1, N + 2, N + 3, \dots, 2N$$

の 2 グループに分ける。そして、それぞれ重心を計算し、その間の距離を求める。そののち、group1 と group2 をそれぞれ 2 つに分割し、それぞれの group 内で同様に重心を計算し、距離を算出する。これより $T = \frac{N}{2}\delta t$ の時間スケールにおける、平均重心位置の移動量が得られる。この操作を繰り返す。



図 3.1: 拘束された分子挙動



図 3.2: 拘束されていない分子挙動

これにより得られる移動量を解析し、主成分分析をおこなうことで目的のデータを得る。



図 3.3: 一部拘束された分子挙動

3.3 解析

データマイニングにより得られたデータから特徴的な結果を示す挙動を抽出し、どの水分子でどの時間に示した挙動なのかを特定し、軌跡をみる。

第4章 データ作成

4.1 シミュレーション

4.1.1 モデル作成

Amber Tools の Leap を用いてシミュレーションモデルを作成した。大きさは一辺が $25\text{\AA}$ の立方体中におよそ 4500 個の分子を配置した。水分子モデルは TIP3P および TIP4PEW を用いた。TIP3P および TIP4PEW のパラメータは表 4.1 に示した通りである。用意したモデルにおける塩分濃度はそれぞれ 0%, 0.43%, 3.5%, 4.8%, 7.7%, 26.4%, 36.4% である。このうち 26.4% は飽和濃度であり、36.4% は飽和濃度を越えたモデルである [25]。これらのモデルを図 4.1–4.12 に示す。ここで図 4.5、4.6 や図 4.11、4.12 ではイオンが水溶媒の外側に配置されているが、計算上は溶媒中に位置取りされている。

4.1.1-a 水分子モデル

水分子モデルは主なものとして TIP3P、TIP4P、SPC/E モデルなどがある。この TIP は Transferable Intermolecular Potential、SPC は Simple Point Charge の頭文字で全て書く原子間の結合長および結合角が拘束された剛体モデルである。この TIP3P や SPC/E は酸素原子と水素原子上に点電荷を持つモデルな

表 4.1: TIP3P および TIP4PEW モデルのパラメータ

	TIP3P	TIP4PEW
$r(\text{OH})$, Å	0.9572	0.9572
HOH, deg	104.52	104.52
$A \times 10^{-3}$, kcal Å ¹² /mol	582.0	656.1
B^6 /mol	595.0	653.5
$q(\text{O})$	-0.834	-0.4844
$q(\text{H})$	+0.417	+1.04844
$r(\text{OM})$	-	0.125

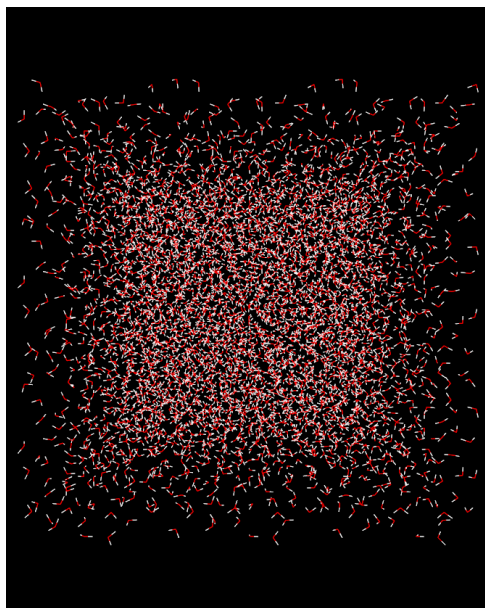


図 4.1: TIP3P を用いた純水モデル

のに対し、TIP4P は結合角 $\angle\text{HOH}$ の二等分線上に点電荷を持つモデルである。

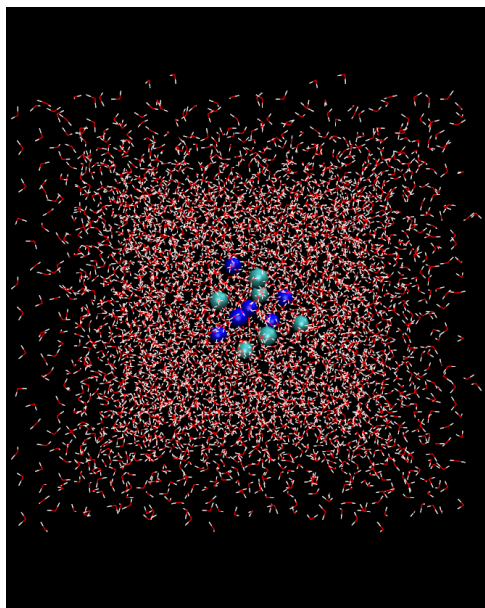


図 4.2: TIP3P を用いた濃度 0.43% の NaCl 水溶液モデル

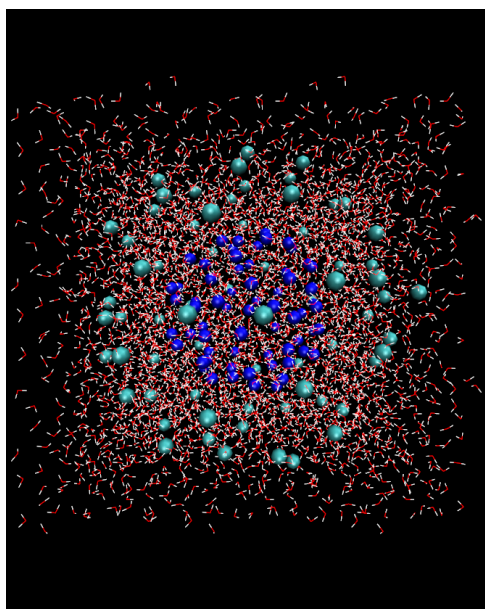


図 4.3: TIP3P を用いた濃度 xx の NaCl 水溶液モデル

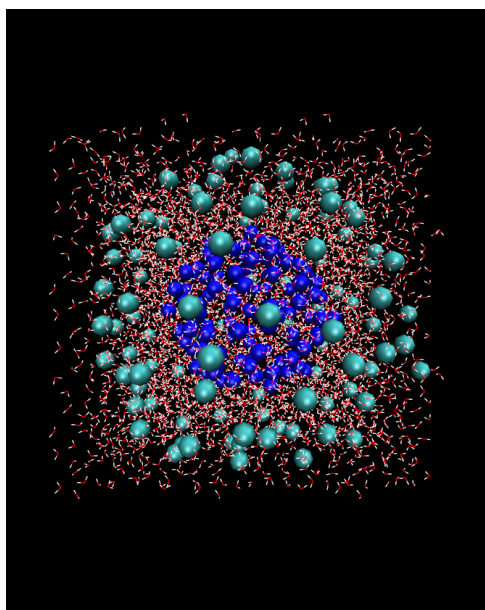


図 4.4: TIP3P を用いた濃度 xx の NaCl 水溶液モデル

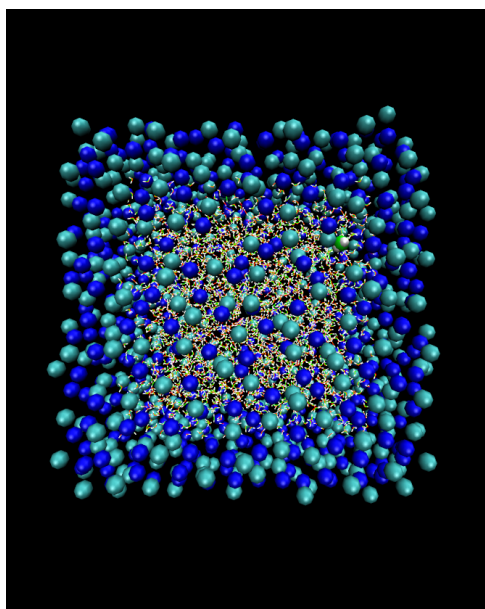


図 4.5: TIP4PEW を用いた飽和濃度の NaCl 水溶液モデル

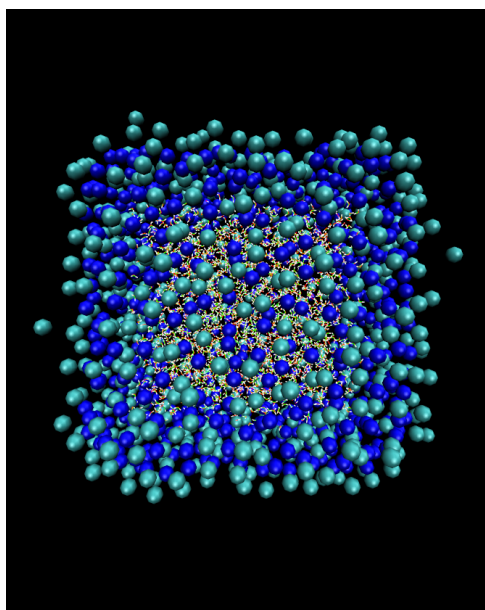


図 4.6: TIP4PEW を用いた濃度 xx の NaCl 水溶液モデル

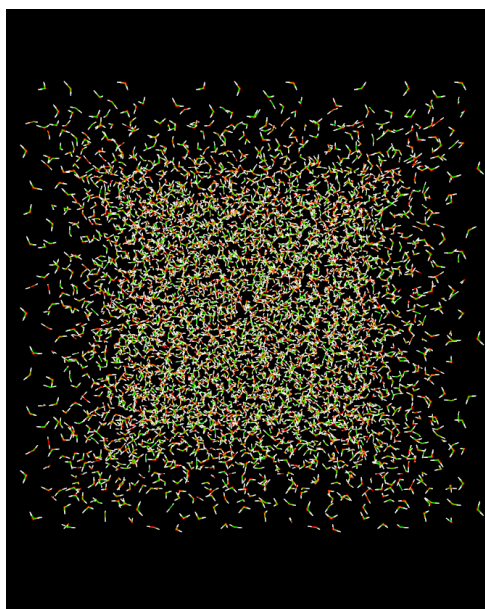


図 4.7: TIP4PEW を用いた純水モデル

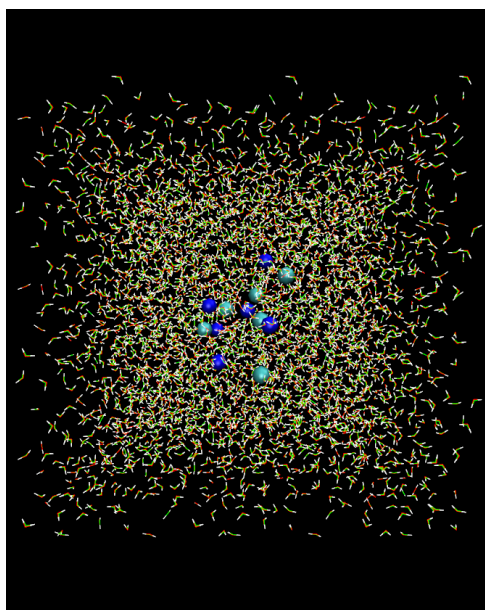


図 4.8: TIP4PEW を用いた濃度 xx の NaCl 水溶液モデル

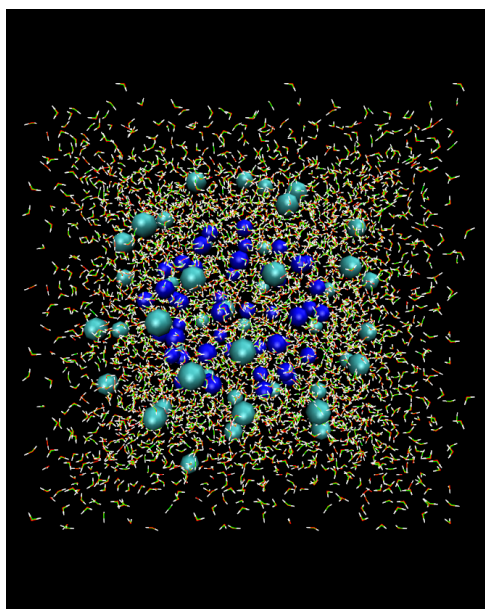


図 4.9: TIP4PEW を用いた濃度 xx の NaCl 水溶液モデル

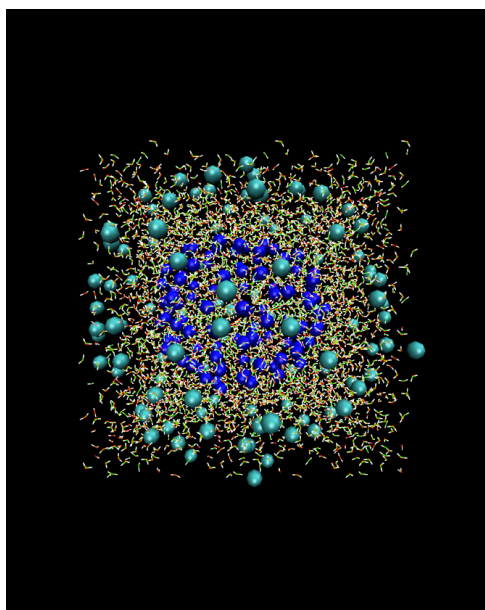


図 4.10: TIP4PEW を用いた濃度 xx の NaCl 水溶液モデル

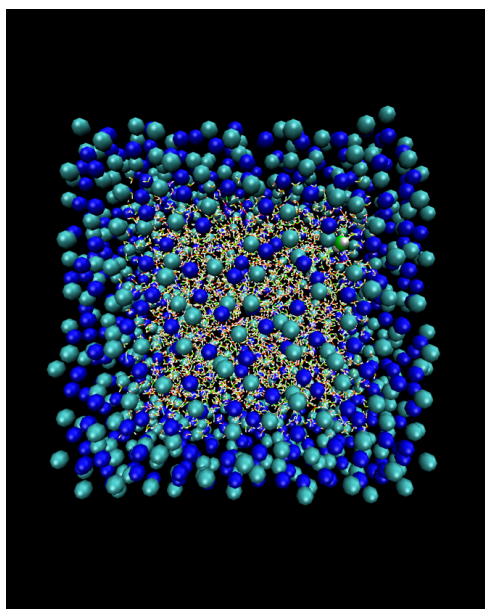


図 4.11: TIP4PEW を用いた飽和濃度の NaCl 水溶液モデル

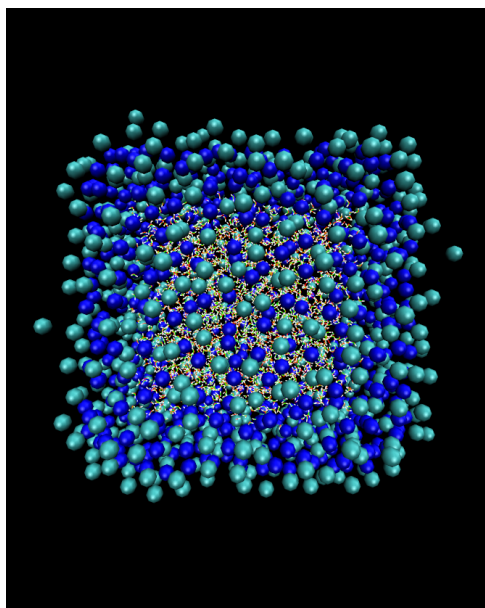


図 4.12: TIP4PEW を用いた濃度 xx の NaCl 水溶液モデル

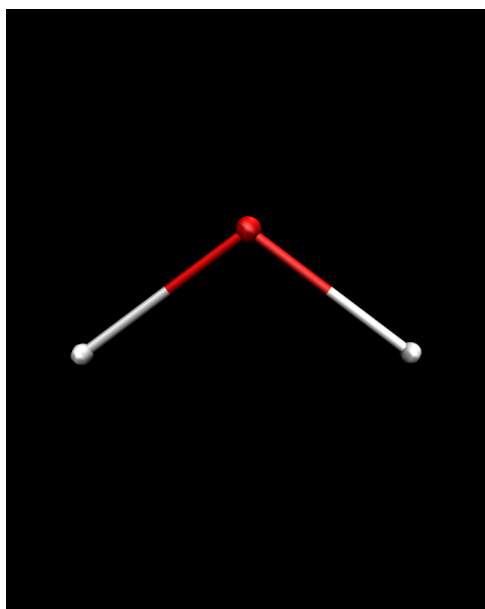


図 4.13: TIP3P モデル

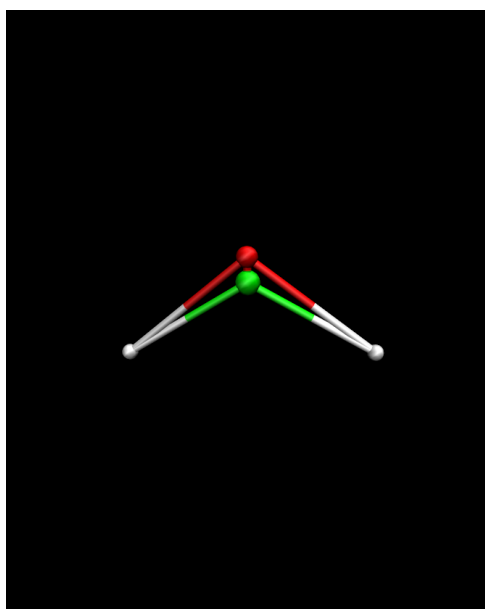


図 4.14: TIP4EW モデル

4.1.2 計算

表 4.1.2 に示した条件でシミュレーションを実行した。これにより、各ステップにおいて分子の位置情報を示す trajectory データを得た。

表 4.2: シミュレーション条件

溶媒モデル	TIP3P および TIP4PEW
Ensemble	NVT ensemble (原子数、体積、温度一定)
温度	300K
総計算時間	200steps = 5ns

4.2 解析

4.2.1 水構造の解析

先行研究において水の構造を解析するためにデータマイニングを用いる手法は用いられている [26]。

水分子の座標データの時間経過による変化を追うことで、水分子の振る舞いを定量化し、解析をおこなった。

解析方法はまず、次のプロセスを繰り返す。

1. 水分子の軌道をステップ数で等分する。
2. 等分した軌道の重心を求める。
3. 重心間距離を求める。

これを一つの分子が 64 等分され、重心間距離が求めるまで繰り返す。

そして、ここでそれぞれの分割ごとに得られた重心間距離に対して次の値を求める。

- 平均値
- 標準偏差
- 最大値
- 最小値

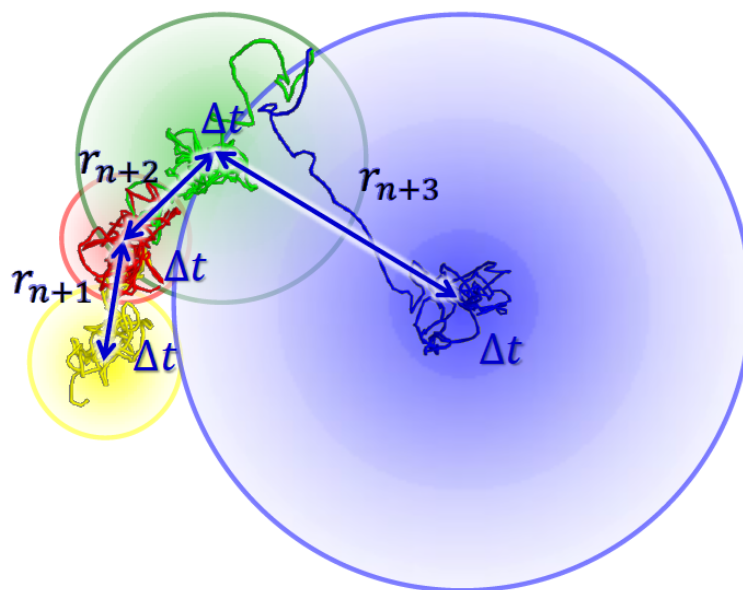


図 4.15: 水分子の軌道の特徴空間に射影し、多次元のベクトルで表現

すると表 4.3 に示したように、上記の値が合計 23 種類得られる。この 23 個の値を用いて水一個の分子挙動を一つの 23 次元のベクトルとして表現した。そのベクトルを主成分分析により解析し、正規化したものをプロットした。

表 4.3: 水分子の分割することによって得られる値の数

n 等分数	2	4	8	16	32	64
重心間距離数	1	3	7	15	31	63
1 フレームあたりのステップ数	512	256	128	64	32	16
得られる値の数	3	4	4	4	4	4

第5章 結果と考察

5.1 NaCl 濃度の影響

NaCl 濃度 0%、0.431%、1.64%、4.74%、32.3%、40.2%の解析をおこない、それぞれの NaCl 水溶液を純水と比較したものを図 5.7 から図 5.10 に示す。これらの結果より NaCl の濃度が上昇するにしたがって特定のモードへ移行することがうかがえる。つまり、濃度が上昇するにしたがって水分子は特定の挙動を示すようになる。

ここで図 5.9 の右下にみられる特徴的な分子の挙動を図 5.13 に示す。この結果は図 5.9 における [0,0] 近傍の典型的な水の挙動である図 5.11 のと比べて挙動が大きく異なる。典型的な水の挙動は

一方、飽和濃度と飽和濃度を超える濃度の NaCl 水溶液の結果を比較したものを図??に示す。ここから飽和濃度以上の濃度では大きな変化は生じないことがわかる。このとき、両者とも水溶液中の濃度はほぼ同じであるものの、40.2%の場合は NaCl の結晶が析出しているという点が異なる。この結果をみると、濃度差は大きいにも関わらず、得られたデータの違いは少ない。

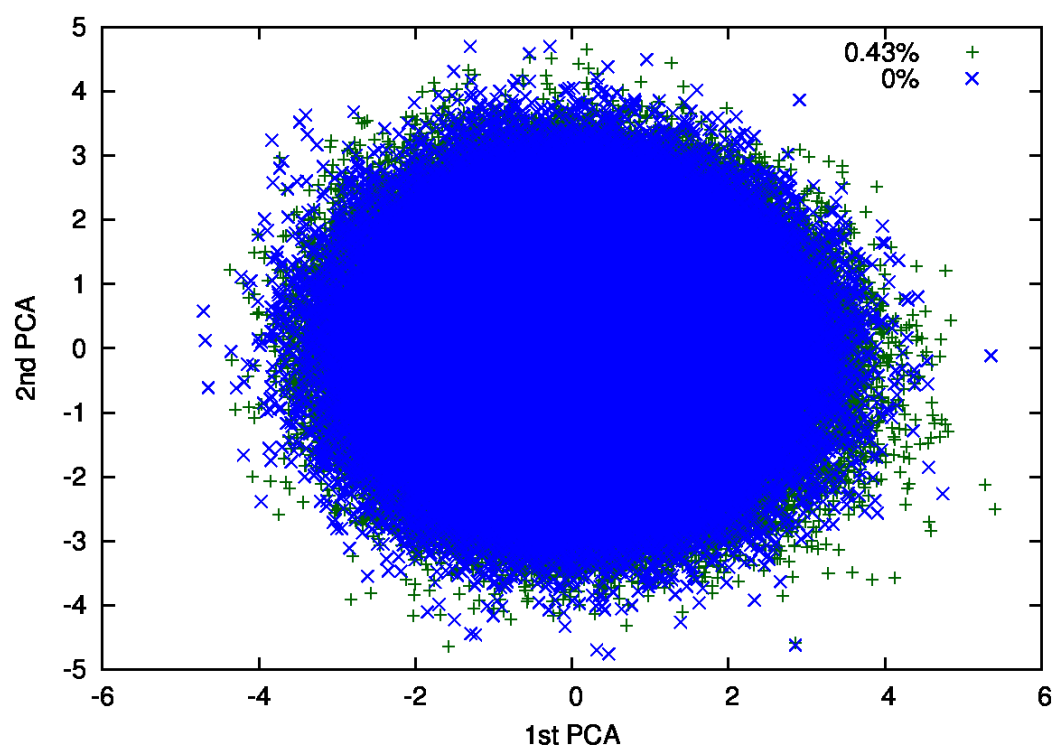
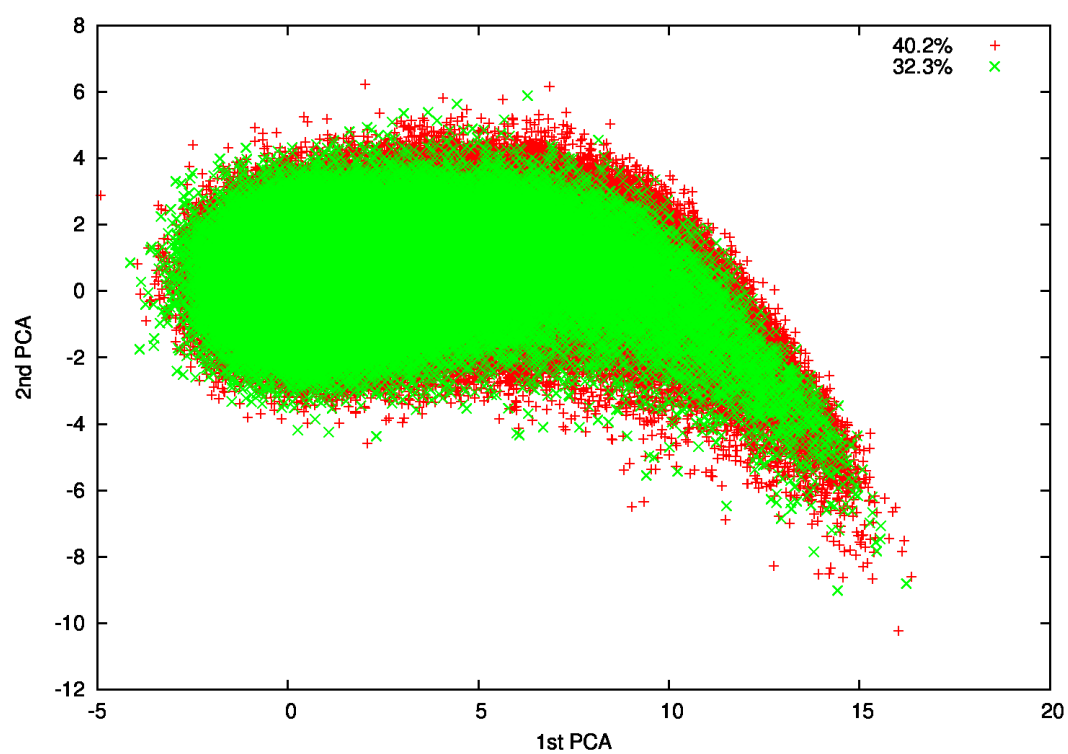


図 5.1: 純水と 0.431%NaCl 水溶液の比較



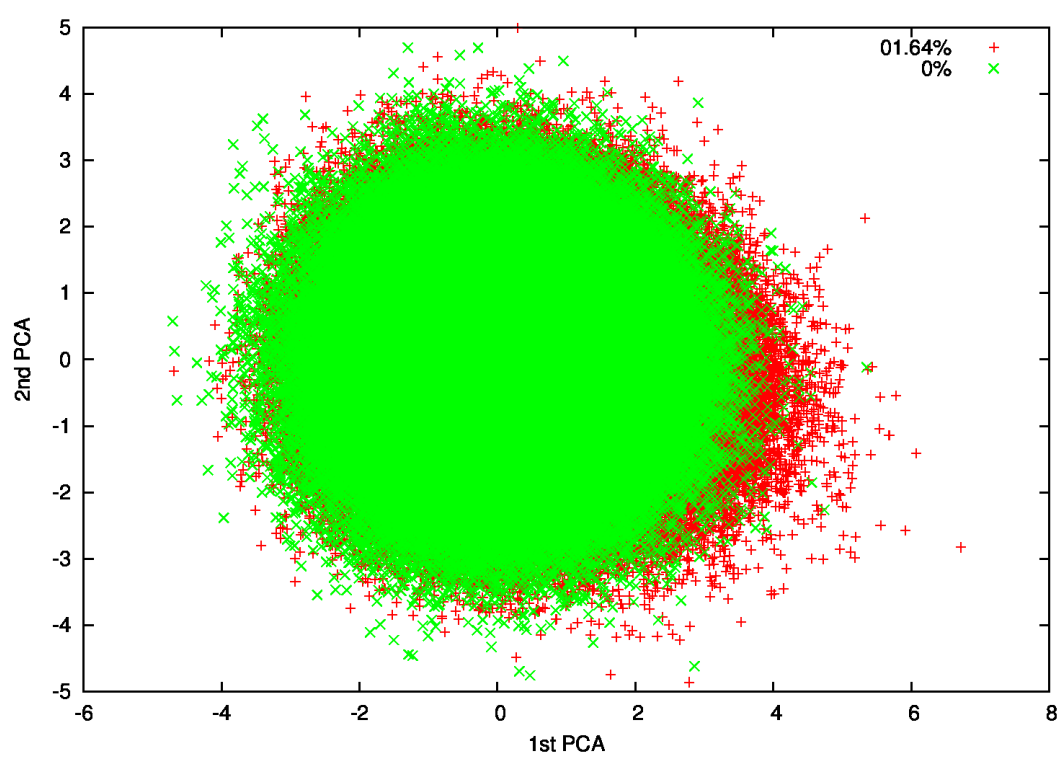


図 5.2: 純水と 1.64%NaCl 水溶液の比較

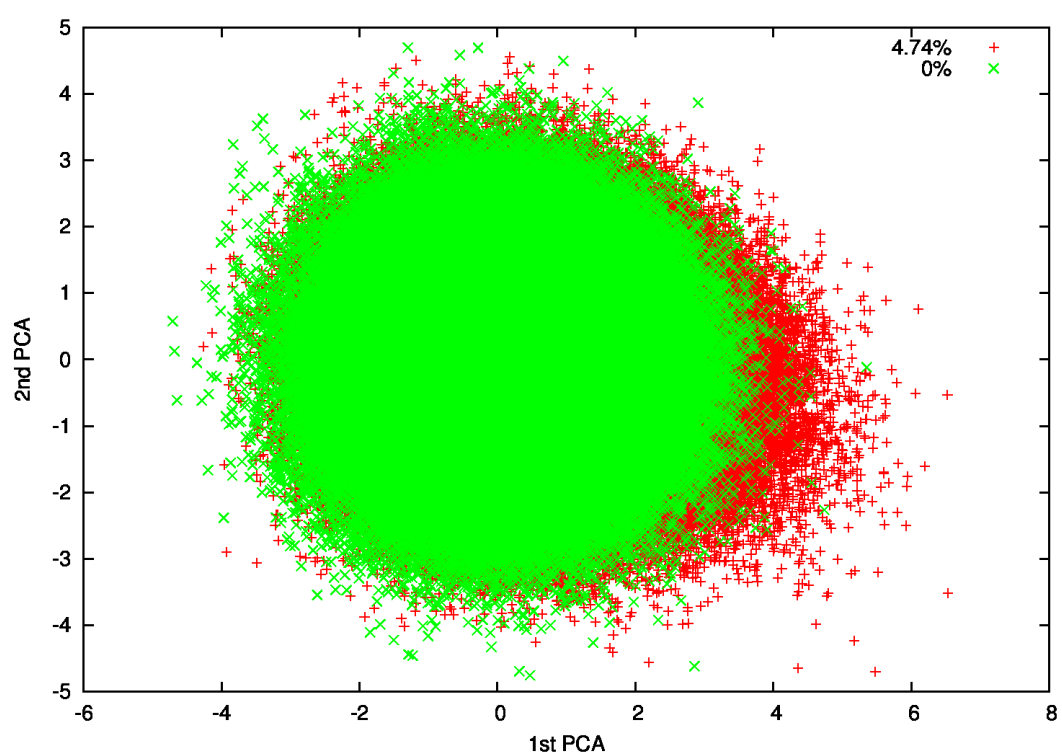


図 5.3: 純水と 4.74%NaCl 水溶液の比較

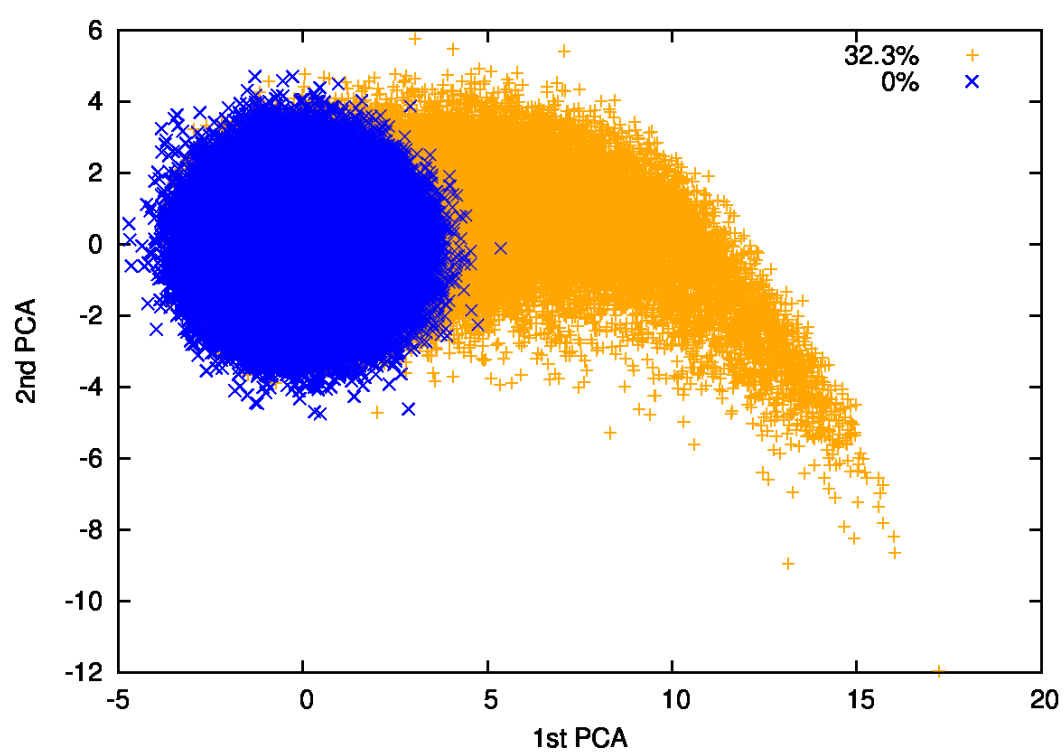


図 5.4: 純水と 32.3%NaCl 水溶液 (飽和濃度) の比較

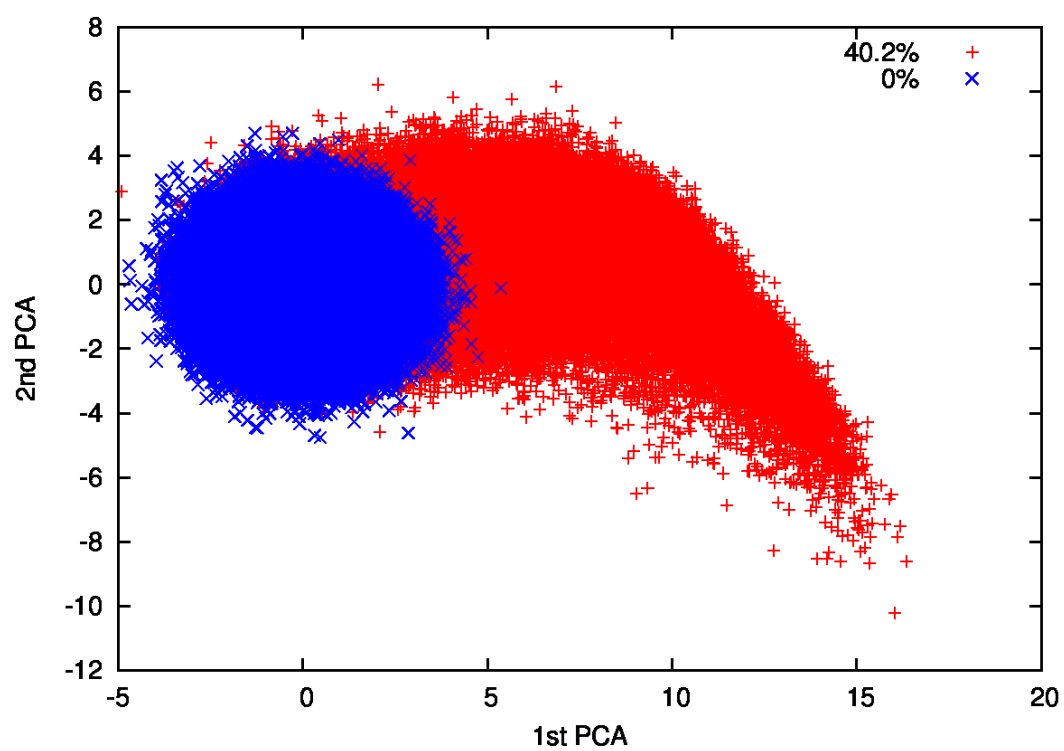


図 5.5: 純水と 40.2%NaCl 水溶液 (飽和濃度を超える濃度) の比較

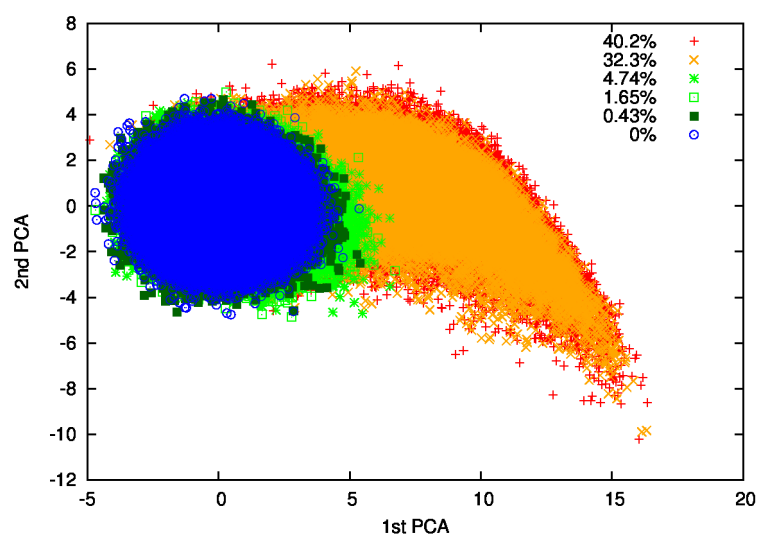


図 5.6: TIP3P モデルのまとめ

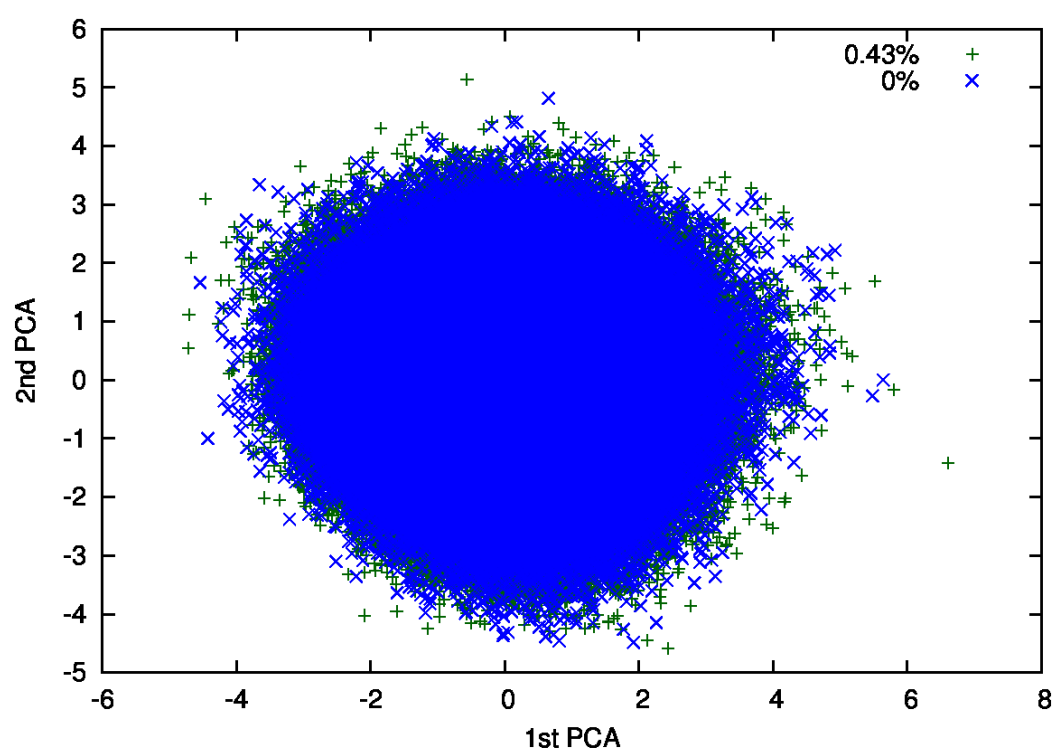


図 5.7: 純水と 0.431%NaCl 水溶液の比較

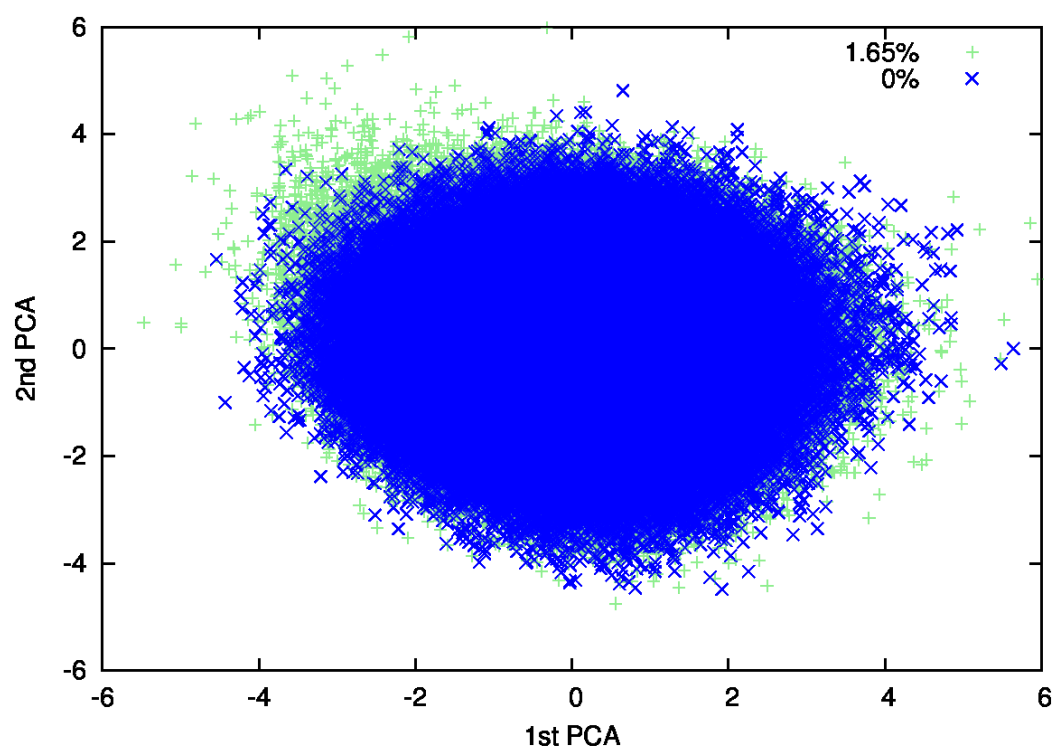


図 5.8: 純水と 1.64%NaCl 水溶液の比較

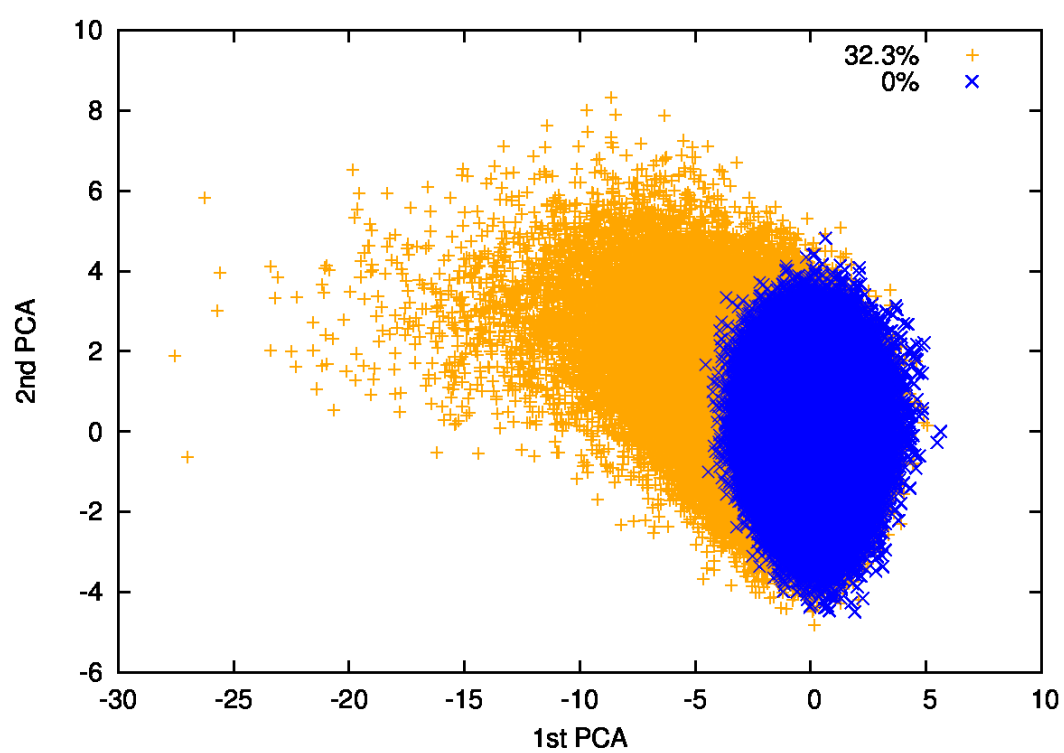


図 5.9: 純水と 32.3%NaCl 水溶液 (飽和濃度) の比較

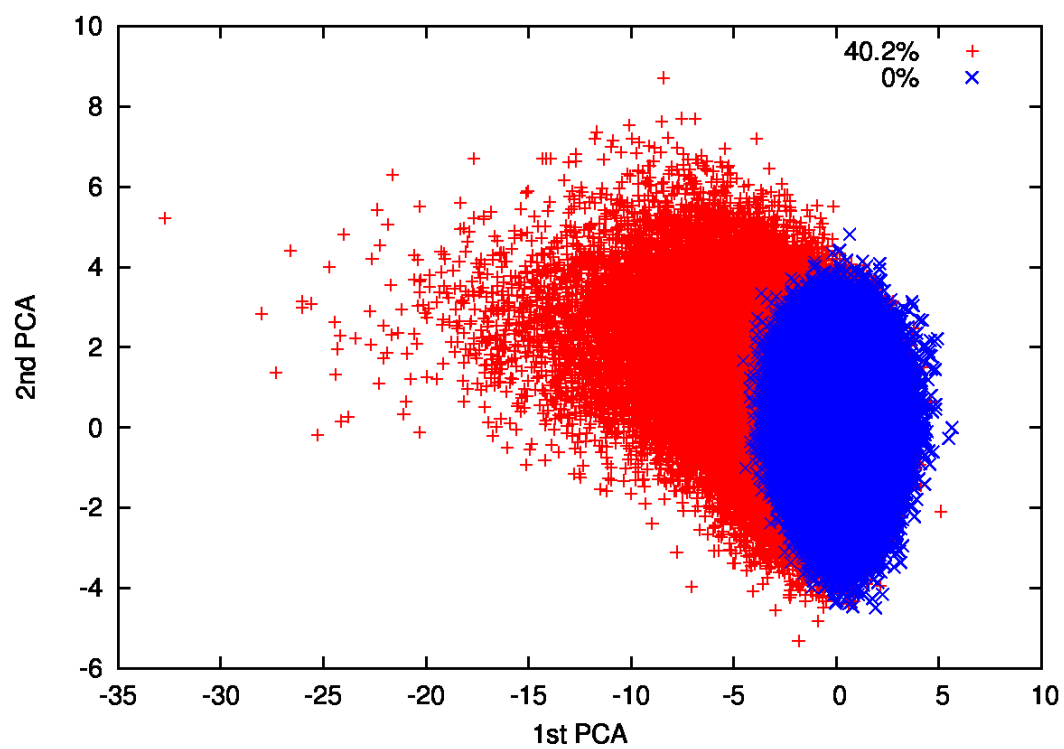


図 5.10: 純水と 40.2%NaCl 水溶液 (飽和濃度を超える濃度) の比較

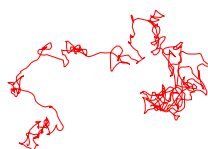


図 5.11: 塩水が示す典型的な水分子の挙動 (表示時間：20ps、 左右長さ：9Å)

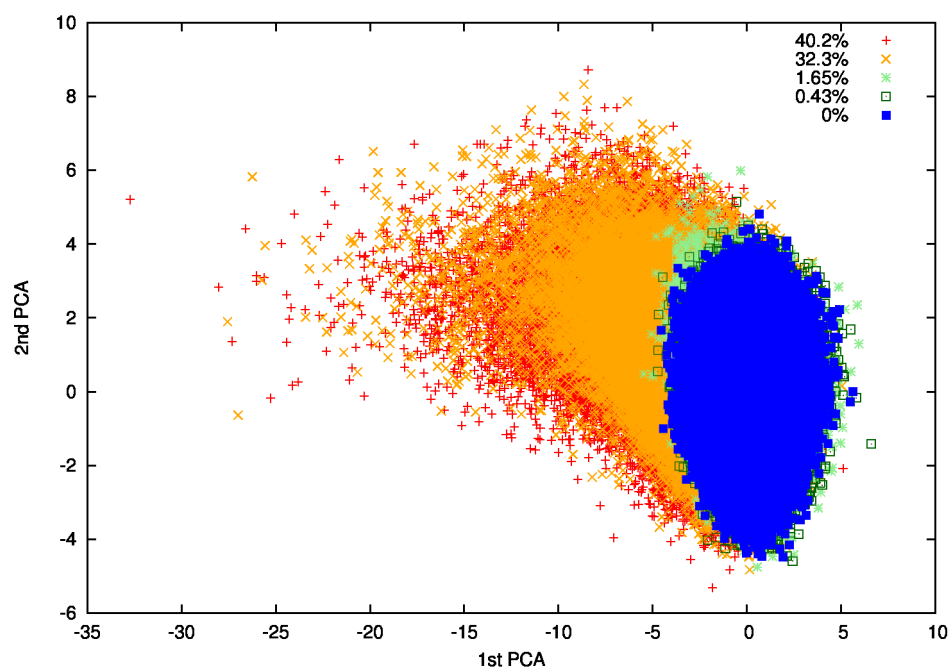


図 5.12: TIP4PEW モデルのまとめ



図 5.13: 飽和濃度食塩水における特徴的な水分子の挙動 (表示時間：20ps、 左右長さ：1.3Å)

5.2 考察

5.2.1 液中濃度

今回の結果から Na および Cl イオン濃度が上昇することで、特定のモードへ移行していくことが見て取れる。これはイオンによって特定の構造を形成しているためではないかと考えられる。その構造を議論するにあたり、はっきりとした構造を形成している高濃度、特に飽和水溶液を議論することは重要である。

今回得られた結果の中で一番興味深かったのが高濃度溶液に現れる図 5.13 のような挙動である。基本的に水の挙動は図 5.11 のように一か所で停滞したり大きく動いたりをランダムに繰り返すが、一か所にとどまり続けるような挙動はしない。そのため、軌跡をみてみると大きく動いていることが確認できる。一方、飽和濃度水溶液に現れる特徴的な軌道は極めて狭い範囲のみに現れる。このときの移動範囲の大きさは最大で $1.3\text{\AA}$ である。これは水分子における水素と酸素の結合距離に近い値である。つまり、この挙動を示すときには結合距離程度しか動かない、おそらく回転運動による動きのみの挙動だろう。

なぜ飽和水溶液濃度の場合にこのような特徴的な挙動、回転運動のみで並行移動がないのだろうか。おそらく、2.2.1 や 2.2.1-a で述べたイオン、特に正の水和をする Na イオンの周囲に水のクラスタが形成されるためだと考えられる。しかし、Na イオン単体であればそれ自体が溶液中を動くため、クラスタを形成する水分子も同様に動くはずである。それが現れなかったのは液中にイオンのネットワークが形成されていたためだと考えられる。この挙動が確認された溶液は飽和水溶液であり、極めて短時間ならば局所的に結晶のようにネットワークを形成する核が生成される可能性がある。それにより周囲に水のクラスタを形成しつつ結晶が生成された構造、結晶水を持つ塩である水和物として存在していることが考えられる。水和物として知られているものは硫酸銅 (II) 五水和物 $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ やヨウ化コバルト (II) 六水和物 $\text{CoI}_2 \cdot 6\text{H}_2\text{O}$ などが挙げられるが、NaCl の水和物というものはあまり聞かない。それというのも、NaCl の水和物は限られた条件でのみ生成されるものであり、室温における飽和水溶液に現れるものではないためだ [27]。それに関らず、このような結果が得られたのは、局所的かつ瞬間的には水和物となりえることを示している。

ここで興味深いことは NaCl 水和物は図 5.14 からわかるように、飽和濃度近傍の水溶液が融点近傍の時に現れる化合物である。それに対して今回の温度は室温であり、水和物は生成されない温度である。この様子をシミュレーションの様子を調べてみると図 5.15、5.16 の結果が得られた。これにより多くのイオンが結晶をつくっており、特徴的な振る舞いを示す水分子はその結晶内に存在することがわかる。つまり、水和物を生成している可能性を示す。

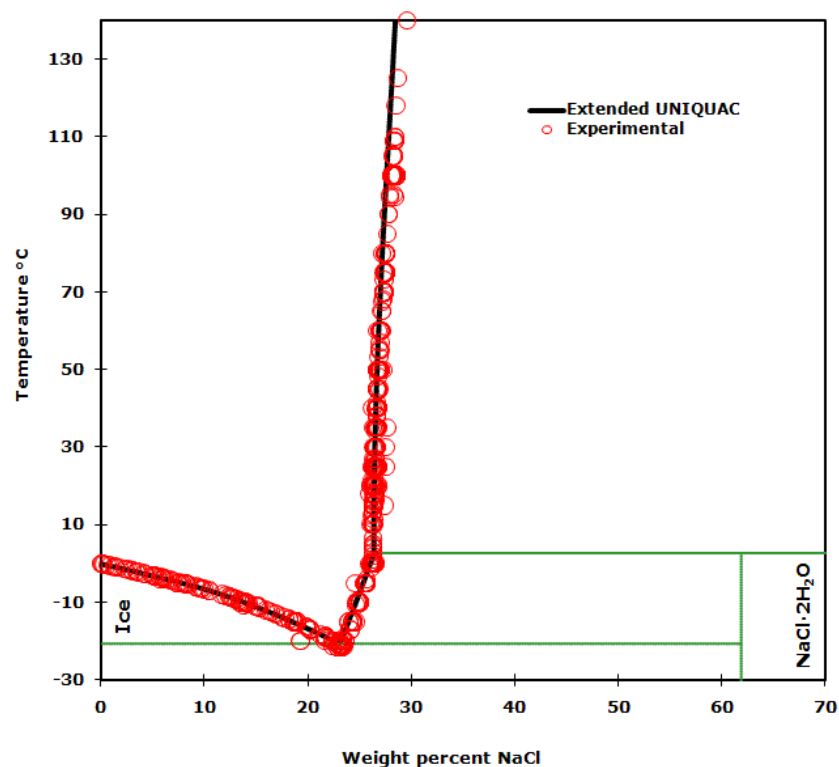


図 5.14: 塩の相図

図 5.7、5.8、5.3、5.9、5.10 からわかるように NaCl 濃度が上昇するにしたがって特定のモードを示すようになる。これは濃度が上昇することで水の構造に Na イオンや Cl イオンが影響を与えているためだと考えられる。なぜなら 2.2.1 や 2.2.1-a で述べたように水分子の挙動はイオンによって大きく影響を受け、イオン同士も影響を与え合うためである。そのためイオンにより純水における水の構造に変化が表れる。イオンの与える影響は濃度が低いうちは小さいもの、あるモードへ移行していくことから含有量が増えるにしたがって特徴的な挙動を示す。

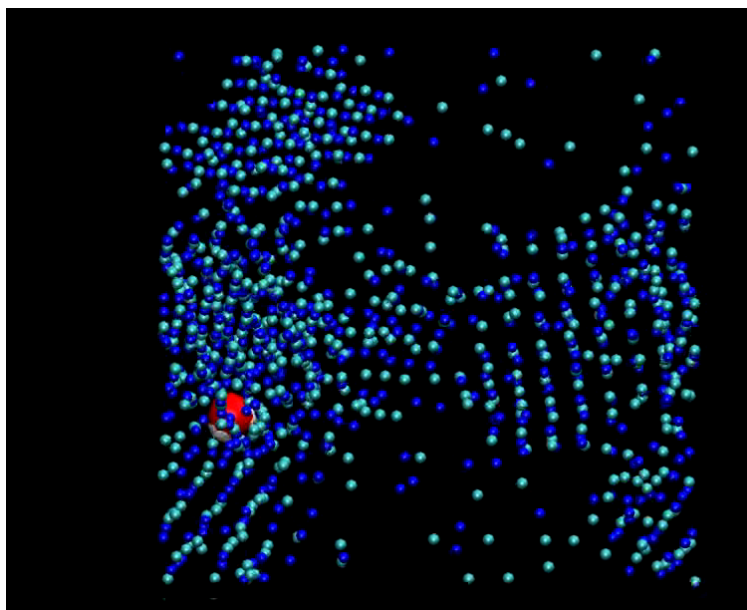


図 5.15: 飽和濃度を超える水溶液中の特徴的な水分子とイオン

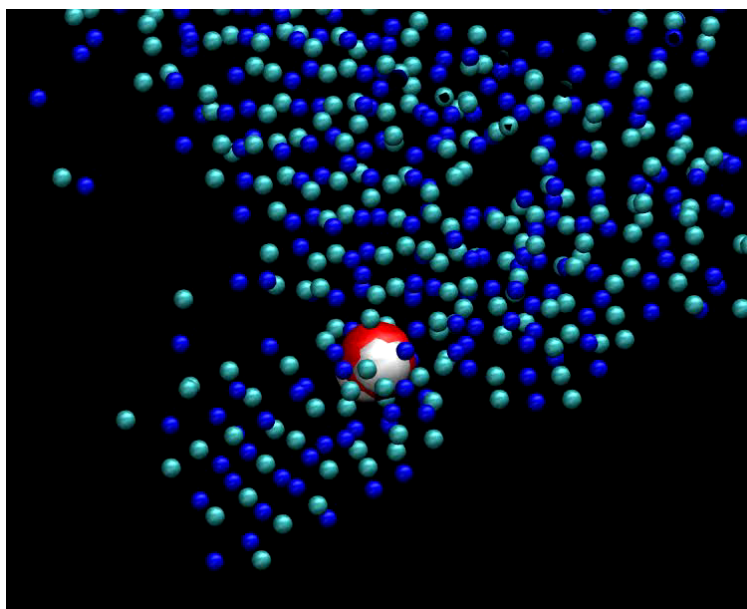


図 5.16: 飽和濃度を超える水溶液中の特徴的な水分子とイオン（拡大図）

5.2.2 飽和濃度以上の水溶液

図??より濃度が大きく異なるにも関わらず、似たモードを示している。これは NaCl 濃度 40.2% の場合には NaCl が結晶として析出し、溶液の濃度が飽和水溶液と同じ状態になるためだからと考えられる。もちろん 2.2.2 で述べたように、結晶が析出したことで結晶界面との相互作用により独特な特性を示すことが考えられる。しかし、今回の結果からその影響を見て取ることは難しい。おそらく結晶は析出しているものの、シミュレーション時間は現実時間に対して短いことから大きな組織をつくるに至らなかったためだと考える。また、この結果から飽和濃度を超える塩水の表現ができたといえる。

第6章 結論

6.1 今後の課題

なにより今回の研究では帰納法的アプローチの有意性を示し、物理的興味深い結果を得ることができたが、データマイニングを用いた解析にまで至ることができなかった。そのため、データマイニングを用いてさらなる水の構造の解明へ研究を進めていく必要がある。

また、今回の結果から NaCl イオン周辺とクラスタ化している水分子への関連性が明らかになった。そのため、今後はイオンと水分子の動的関係を明らかにするとともに、イオン同士のネットワーク構造を調査することが重要となる。それにより、NaCl の水和物に対する新たな理解が進む可能性を持つ。

しかし一方、今回の結果ではデータが少なく、イオン濃度変化による詳細な議論に至ることができなかった。今後の課題としては高濃度と低濃度の間の濃度のデータを得ることでより塩水中の水分子挙動について理解が進むと考える。また、イオンの挙動も解析すること塩水全体の分子が作り出す構造に関して理解が得られるだろう。

6.2 まとめ

塩水は自然界に多く存在し、私たちの生活からも欠かせない重要な溶液のひとつである。そのため昔から多くの科学者が研究の対象としてきた。そのなかで、計算機の発達によりより大きな系、長時間のシミュレーションによる研究が進められてきた。しかし、得られるデータが巨大なものとなっているにも関わらず多くが従来の解析手法を用いている。

そこで本研究はデータマイニングの手法を用いることで、巨大データに対して帰納法的アプローチをおこない、従来の手法では得られなかった知識を発見することを目的とした。その結果、系全体の挙動を解析することで、イオン周囲に水のクラスタが形成され、イオンによるネットワーク構造が形成されている可能性を示すことができた。これは従来室温は生成されないと考えられてきた NaCl の水和物である可能性が

ある。これにより、塩水の性質に関して新たな可能性を示唆することができた。以上のことから帰納法的アプローチによる解析の有効性を示すことができた。

今後は、この結果を生かし、データマイニングを用いてさらに動的構造を調査し、検証していくことで塩水の特徴のさらなる解明につなげていきたい。

謝辞

本研究を遂行し学位論文をまとめるに当たり、主指導教官として多くのご支援とご指導を賜りました、北陸先端科学技術大学院大学知識科学研究科知識科学専攻 DAM HIEU CHI 准教授 には深く感謝しております。時に応じて、厳しくご指導頂いたこと、また優しく励まして下さったことを通して、私自身の至らなさを実感することが出来たことは今後の努力の糧になる物であります。また同専攻 Ho Tu Bao 教授には副指導員としてご助言を戴くとともに本論文の細部にわたりご指導を戴いたことに感謝しております。本研究に取り組むに当たり、日々の研究だけでなく多くのご指導をしていただいた同大学マテリアルサイエンス研究科マテリアルサイエンス専攻 水上卓 助教授並びに ERATO 下田ナノ液体プロセスプロジェクト研究員 杉山歩さんには深謝しております。また2年間の大学院生活においてやってこれたのは DAM 研究室および近隣研究室の方々が様々な面から支えてくれたためです。ありがとうございます。

付 録 A Amber

Amber(Assisted **m**odel **b**uilding with **e**nery **r**efinement) は、モデリングおよび分子力学と動力学計算シミュレーションを行うために開発されたプログラムのひとつであり [28]、タンパク質や核酸の構造解析などに用いられている。

本プログラムはカリフォルニア大学サンフランシスコの故 Peter Kollman の手により開発が始められたものであり、現在は Scripps 研究所の David Case が中心となって改良が続けられている。本プログラムの最新版は 2010 年 4 月 25 日にリリースされた Amber11 となっており [29]、WEB 上で公開されている。本研究ではこの最新版を使用してシミュレーションを実行した。

A.1 leap

分子シミュレーションを実行するためにはどのような系のシミュレーションを実行するかモデルを作成する必要がある。leap は Amber でシミュレーションを実行するためのソフトである。このソフトに力場を読み込ませ、シミュレーションモデルのサイズを決め、分子や原子、イオン、溶媒などを配置する。これを用いて計算を実行する。

A.2 sander

分子シミュレーションを実行するプログラム。他のサブ的プログラムを用いてモデルや条件設定など input ファイルを作成し、このプログラムで読み込み、計算を実行することでデータを得る。

参考文献

- [1] 日本金属学会. 金属物理化学 (金属化学入門シリーズ). 日本金属学会, 5 1996.
- [2] 橋本佑介, 中嶋明, 奥村哲也, 専徳博文, 森高秀四郎. J22 3 機素潤滑設計 (トライボロジー). 日本機械学会九州支部講演論文集, Vol. 2009, No. 62, pp. 295–296, 2009-03-18.
- [3] 伏木亨. 味覚と嗜好のサイエンス [京大人気講義シリーズ]. 丸善, 4 2008.
- [4] 稲田仁, 富永真琴. 温度・辛み・酸味センサー:trp チャネルの多様性 (特集感覚センサー—生物が外界環境を感じる能力を探る). 細胞工学, Vol. 26, No. 8, pp. 878–882, 2007-08.
- [5] Maik Behrens and Wolfgang Meyerhof. Gustatory and extragustatory functions of mammalian taste receptors. *Physiology & Behavior*, Vol. 105, No. 1, pp. 4–13, 2011.
- [6] Grace Q Zhao, Yifeng Zhang, Mark A Hoon, Jayaram Chandrashekar, Isolde Erlenbach, Nicholas J P Ryba, and Charles S Zuker. The receptors for mammalian sweet and umami taste university of california at san diego. *October*, Vol. 115, No. 3, pp. 255–266, 2003.
- [7] 都甲潔. 味覚センサ-. 日本化学会誌 : 化学と工業化学 = Journal of the Chemical Society of Japan : chemistry and industrial chemistry, Vol. 1995, No. 5, pp. 334–342, 1995-05-10.
- [8] 高野仁, 北村由香, 秋谷昌宏. Lb 膜味覚センサにおける識別能力向上の検討. 電子情報通信学会技術研究報告. OME, 有機エレクトロニクス, Vol. 100, No. 556, pp. 53–58, 2001-01-11.
- [9] Donald Kennedy and Colin Norman. What don't we know? *Science*, Vol. 309, pp. 78–102, 2005.
- [10] 上平恒. 水の分子工学. 講談社, 4 1998.
- [11] 上平恒. 水とはなにか 新装版 (ブルーバックス). 講談社, 新装, 7 2009.
- [12] 谷口尚司, 八木順一郎. 材料工学のための移動現象論. 東北大学出版会, 1 2002.

- [13] 飯坂剛一, 福島信弘, 和泉研二. 水溶液中における塩化ナトリウム結晶表面の分子動力学シミュレーション : 基礎 (理論)iii. 日本結晶成長学会誌, Vol. 21, No. 3, p. 287, 1994-07-05.
- [14] Jean-Pierre Hansen and Ian R. McDonald. *Theory of Simple Liquids, Second Edition*. Academic Press, 2 edition, 2 1990.
- [15] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford Univ Pr on Demand, reprint edition, 6 1989.
- [16] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)*. Academic Press, 2 edition, 10 2001.
- [17] 大塚亮, 片岡洋右, Otsuka Ryo, Kataoka Yosuke. NaCl 水溶液中における相転移の分子動力学シミュレーション. 2011-06-01.
- [18] N Matubayasi and M Nakahara. Reversible molecular dynamics for rigid bodies and hybrid monte carlo. *JOURNAL OF CHEMICAL PHYSICS*, Vol. 110, No. 7, pp. 3291–3301, FEB 15 1999.
- [19] TF Miller, M Eleftheriou, P Pattnaik, A Ndirango, D Newns, and GJ Martyna. Symplectic quaternion scheme for biophysical molecular dynamics. *JOURNAL OF CHEMICAL PHYSICS*, Vol. 116, No. 20, pp. 8649–8659, MAY 22 2002.
- [20] 杉山公造, 下嶋篤, 梅本勝博, 橋本敬, 永田晃也. ナレッジサイエンス - 知を再編する 81 のキーワード. 近代科学社, 改訂増補, 3 2008.
- [21] 元田浩, 山口高平, 津本周作, 沼尾正行. データマイニングの基礎 (IT Text). オーム社, 12 2006.
- [22] ピーターエイドリアン, ドルフザンティンジ. データマイニング. 共立出版, 6 1998.
- [23] 豊田秀樹. データマイニング入門. 東京図書, 12 2008.
- [24] 岡嶋裕史. 数式を使わないデータマイニング入門 隠れた法則を発見する (光文社新書). 光文社, 5 2006.
- [25] 国立天文台 (編). 理科年表 平成 21 年. 丸善, 平成 21 年, 11 2008.
- [26] Hieu-Chi Dam, Tu Bao Ho, and Ayumu Sugiyama. Simulation-based data mining solution to the structure of water surrounding proteins. In *IJCAI*, pp. 2424–2429, 2011.

[27] The sodium chloride - water binary system.

[28] 長岡正隆. すぐできる分子シミュレーションビギナーズマニュアル (KS 化学専門書). 講談社, 11 2008.

[29] Amber home page. <http://ambermd.org/>.